Building Recommendation Systems

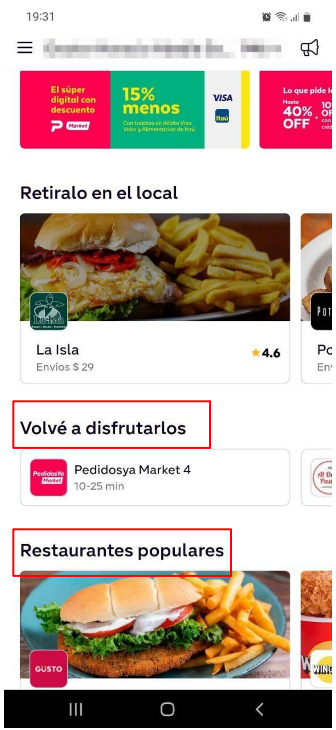
Cesar Reyes

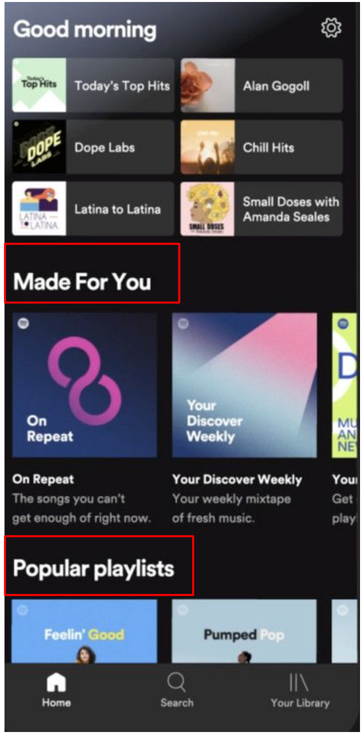
Principal Data Scientist

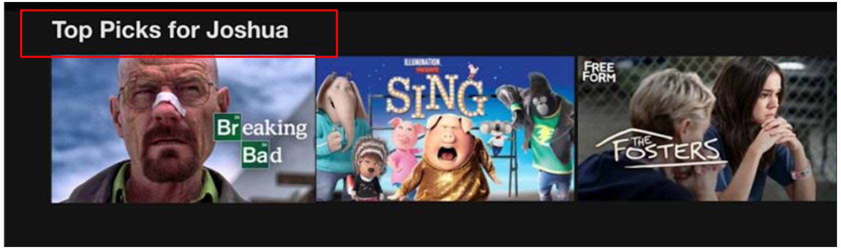
PedidosYa

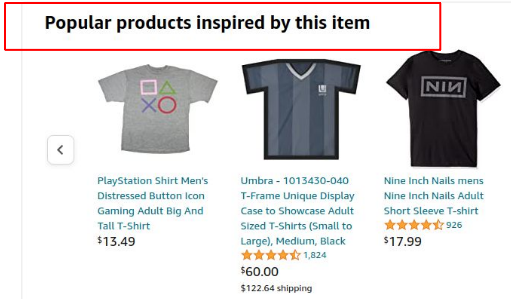
What's a RecSys

We are interacting with RS every day



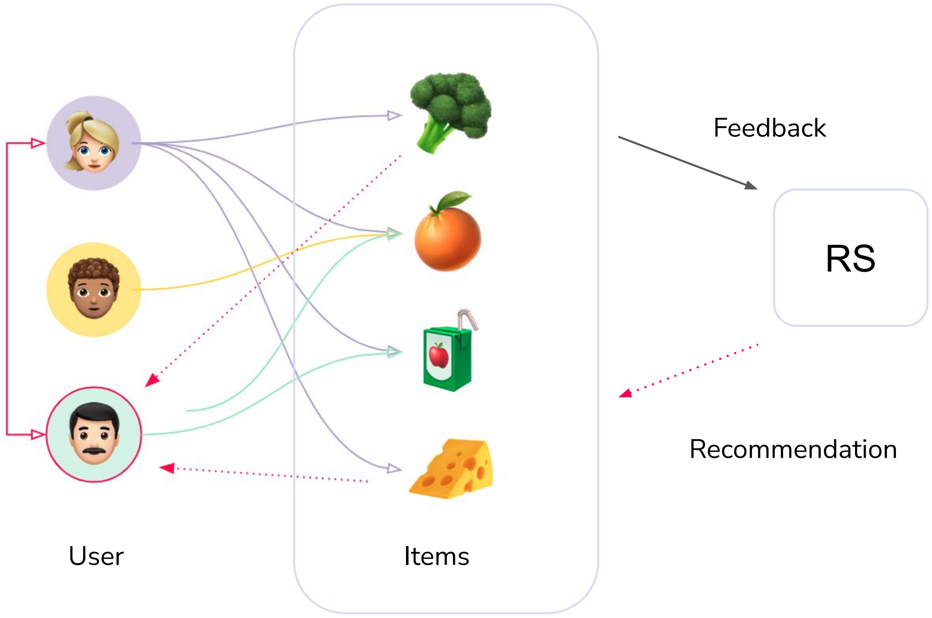






“Recommender systems help solve information overload by helping users find **relevant products** from a **wide range of selections** by providing personalized content.”

What’s a RecSys?



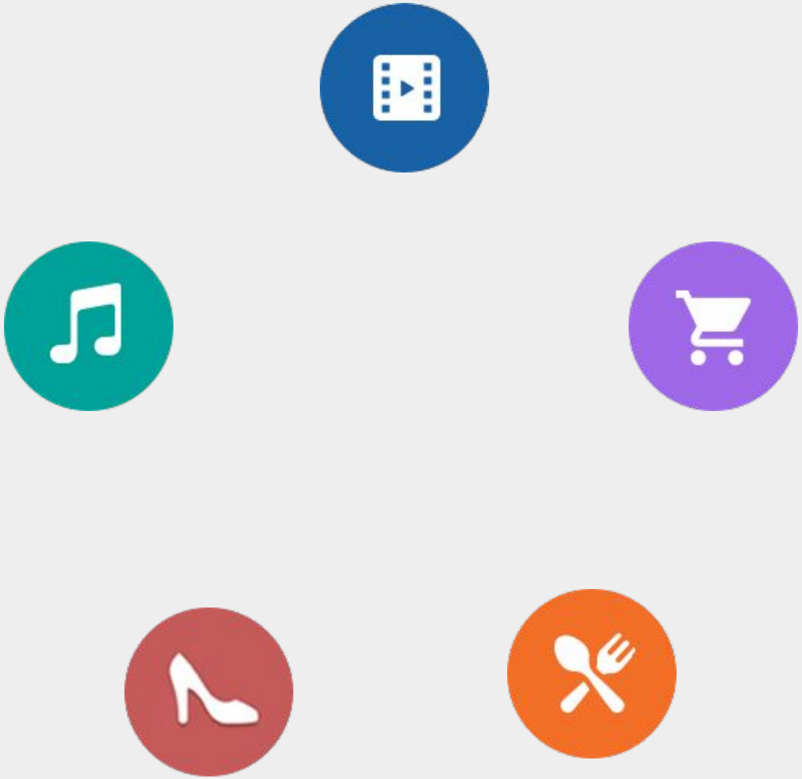
A recsys calculates and provides **relevant content to the user based on knowledge of the user, content, and interaction** between the user and the item

Component of a RS

Components

Taxonomy

- Domain
- Purpose
- Context
- Personalization level
- Privacity and trustworthiness
- Interface
- Algorithms



Taxonomy

- Domain
- **Purpose**
- Context
- Personalization level
- Privacity and trustworthiness
- Interface
- Algorithms

+ CVR?

+- time in app?

+ coverage of the sales?

+ experience?

+ retention?

+ active users?

one-time visitor?

delivery service?

paying the subscription?

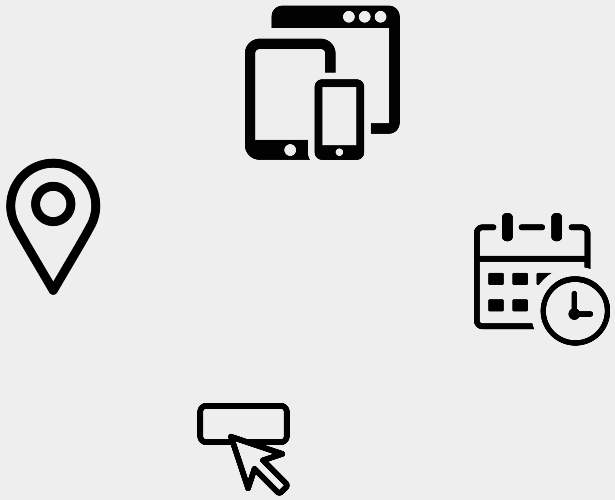
e-commerce or subscription platform?

new or heavy users?.

Components

Taxonomy

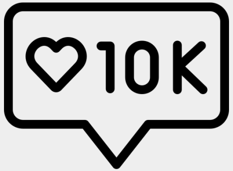
- Domain
- Purpose
- **Context**
- Personalization level
- Privacity and trustworthiness
- Interface
- Algorithms



Components

Taxonomy

- Domain
- Purpose
- Context
- **Personalization level**
- Privacity and trustworthiness
- Interface
- Algorithms



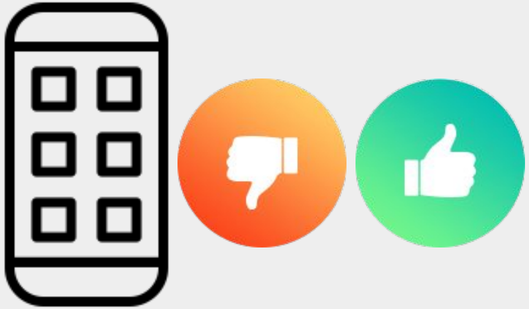
Taxonomy

- Domain
- Purpose
- Context
- Personalization level
- **Privacity and trustworthiness**
- Interface
- Algorithms

Components

Taxonomy

- Domain
- Purpose
- Context
- Personalization level
- Privacity and trustworthiness
- **Interface**
- Algorithms



Components

Taxonomy

- Domain
- Purpose
- Context
- Personalization level
- Privacity and trustworthiness
- Interface
- Algorithms

Content base

Collaborative Filtering

Hibrid

Designing

Think about your current situation: value and impact

Business Goal

Increase retention? CVR? LTV?

How the impact and success of the recommender can be measured?

Look at your problem

Which types of recommendations create value, both for the consumer and the provider of the recommendations?

Is the novelty, popularity, continuation, metadata important?

Interaction

What are the interaction values? Are these available?

How many users? How many items?

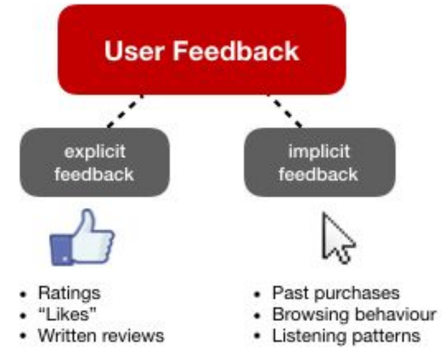
Additional Features

What are the user/item features? Are available online? Are useful for recsys task?

How to collect data - User behavior and how to collect it

Evidence is the data that reveals the user's tastes and preferences. When we talk about collection evidence, we're collection **events** and **behavior** that provide an indicator of the user's tastes.

Generally, two type of feedback are produced by users of a system, Explicit and implicit.



Context

Start simple and iterate

User the previous strategy as baseline

Non-personalize

1. Popular / Novelty
2. Use item metadata

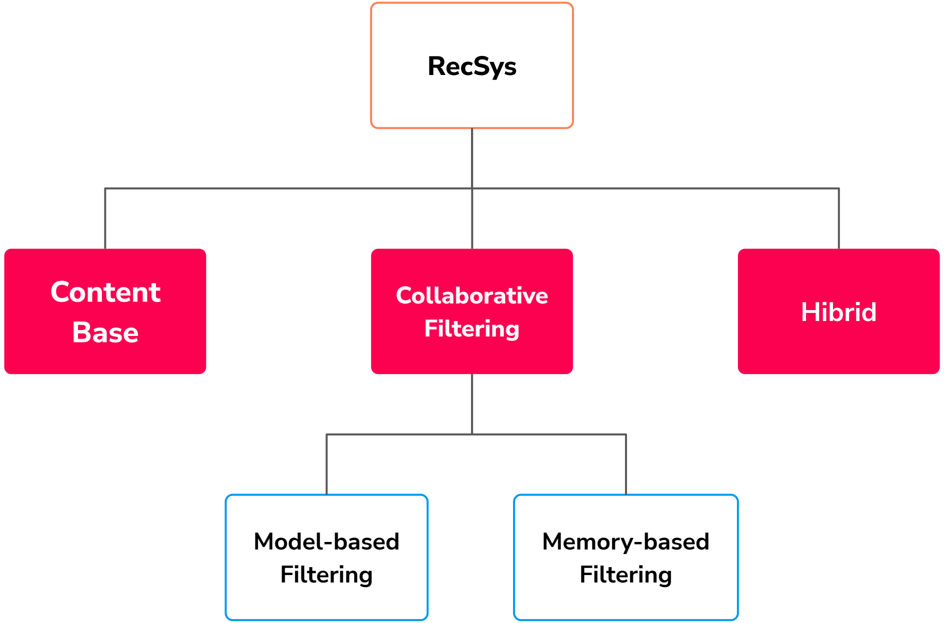
Personalize

3. User and item data
4. User, item and context data

People need to see things they are familiar with to believe that the Recommendation System makes good recommendations. Otherwise, you may think the recommendation system is bad and not interact with it, or worse, you may make fun of it on social media.

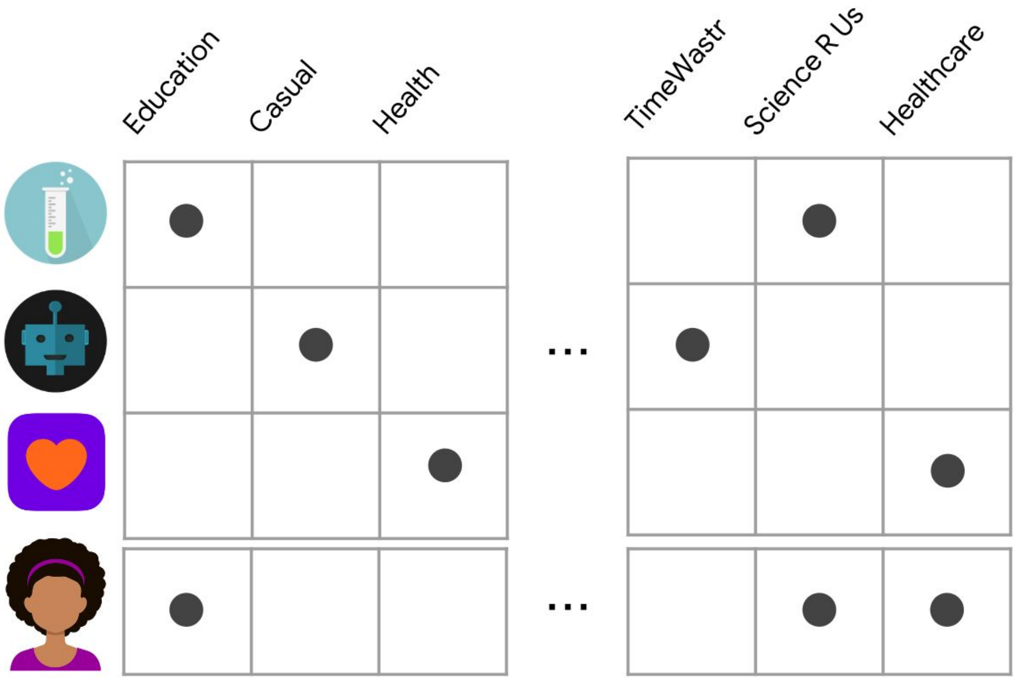
ML Models

Types of Recsys



How do you choose the model that constitute good model?

Content base



KNN

https://developers.google.com/machine-learning/recommendation

Collaborative Filtering - Matrix Factorization

	Harry Potter	The Triplets of Belleville	Shrek	The Dark Knight Rises	Memento
	✓		✓	✓	
		✓			✓
	✓	✓	✓		
				✓	✓

≈

V

.9	-1	1	1	-.9
-.2	-.8	-1	.9	1

U

1	.1
-1	0
.2	-1
.1	1

.88	-1.08	0.9	1.09	-0.8
-0.9	1.0	-1.0	-1.0	0.9
0.38	0.6	1.2	-0.7	-1.18
-0.11	-0.9	-0.9	1.0	0.91

A ~ U x V

SVD, ALS, SGD, WMF

https://developers.google.com/machine-learning/recommendation

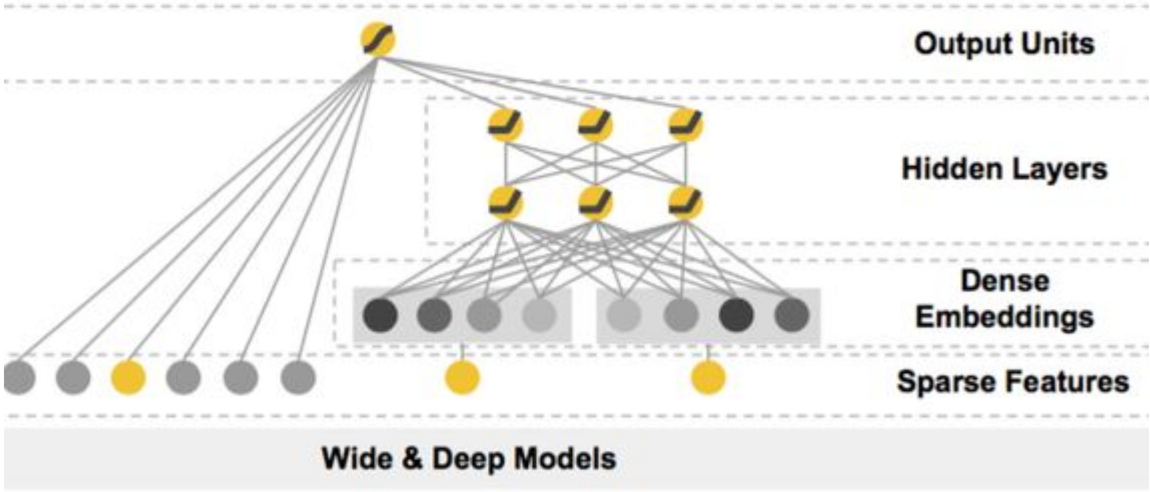
Hybrid - SLIM (Sparse Linear Methods for Top-N Recommender Systems) / FM

Feature vector **x**

	User				Movie					Other Movies rated					Time	Last Movie rated					Target y	
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...		TI	NH	SW	ST	...		
$\mathbf{x}^{(1)}$	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	$y^{(1)}$
$\mathbf{x}^{(2)}$	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	$y^{(2)}$
$\mathbf{x}^{(3)}$	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	$y^{(2)}$
$\mathbf{x}^{(4)}$	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	$y^{(3)}$
$\mathbf{x}^{(5)}$	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	$y^{(4)}$
$\mathbf{x}^{(6)}$	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	$y^{(5)}$
$\mathbf{x}^{(7)}$	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	$y^{(6)}$

http://glaros.dtc.umn.edu/gkhome/node/774

DL - Wide & Deep Learning for Recommender Systems



https://arxiv.org/abs/1606.07792

DL - Neural Collaborative Filtering

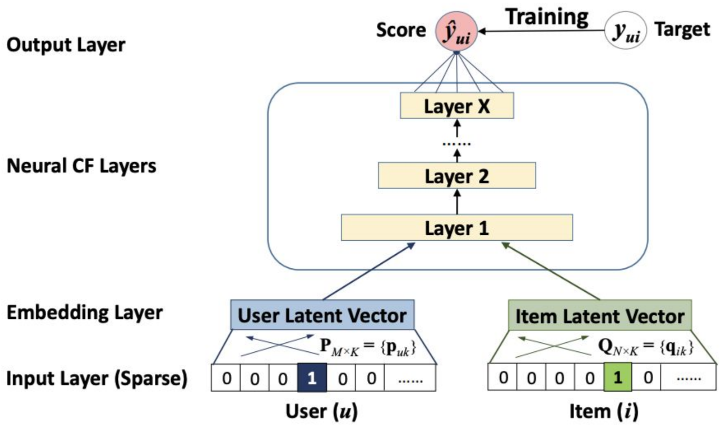


Figure 2: Neural collaborative filtering framework

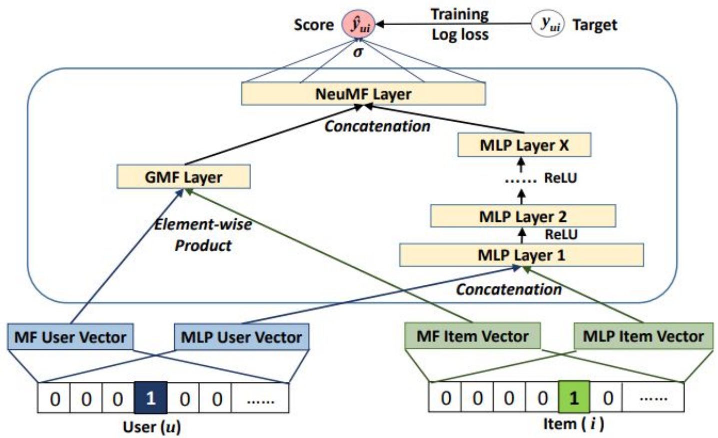


Figure 3: Neural matrix factorization model

https://arxiv.org/abs/1708.05031

DL - DeepFM

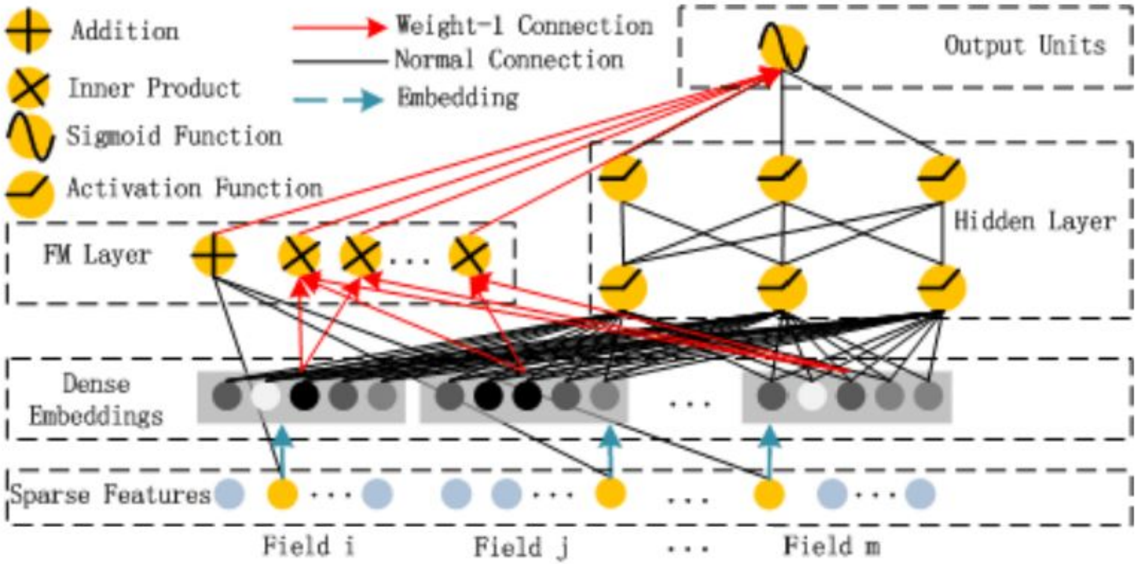
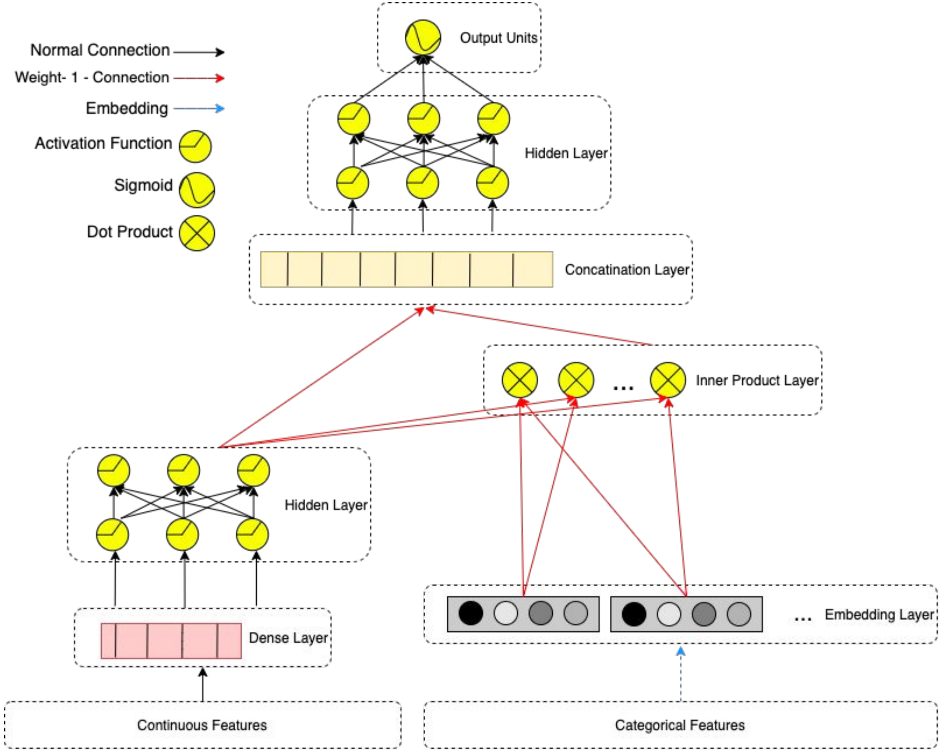


Figure 1: Wide & deep architecture of DeepFM. The wide and deep component share the same input raw feature vector, which enables DeepFM to learn low- and high-order feature interactions simultaneously from the input raw features.

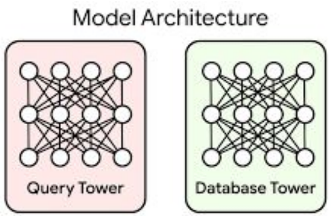
https://arxiv.org/abs/1703.04247

DL - DLRM (Deep Learning Recommendation Model)



http://glaros.dtc.umn.edu/gkhome/node/774

DL - Two tower model

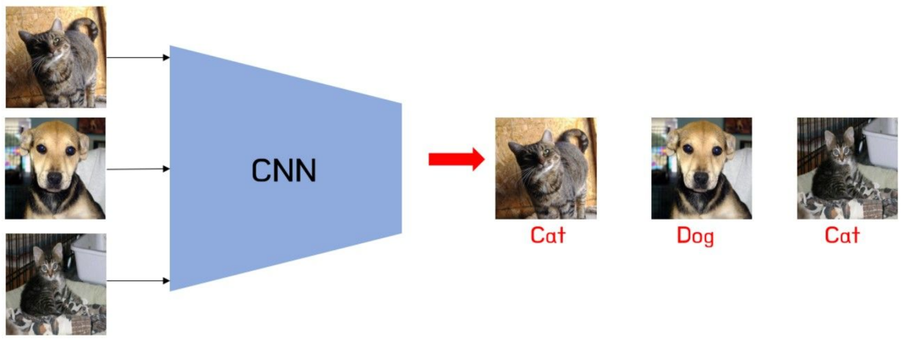


https://research.google/pubs/pub48840/

"In the same way that **word embeddings** revolutionized NLP, **item embeddings** are revolutionizing recommendations"

Evaluating and testing RecSys

How do you know that one model is better than another?



(a) List title: "Fantasy books" (b) List title: "My favorite bo oks in 2019"

Traditional ML problem has a well defined label that can be use to measure offline classification metrics

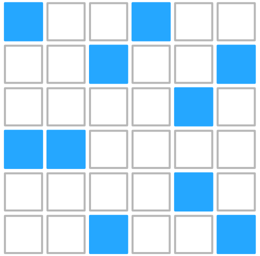
In Recsys the ranking metrics are the key to measure the model performance. The most important item should be at the top of the recommendation

“It is important to emphasize that recommendation often involves solving a surrogate problem and transferring the result to a particular context. A classic example is the assumption that accurately predicting ratings leads to effective movie recommendations [2]. We have found that the choice of this surrogate learning problem has an outsized importance on performance in A/B test”. [Deep Neural Networks for YouTube Recommendations]

How to Split train / test split?

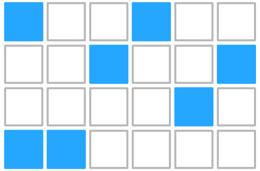
Dataset

Original

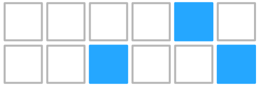


Traditional ML

Train

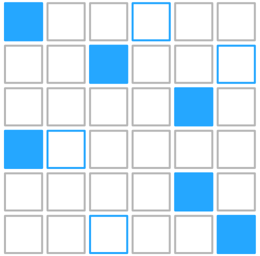


Test

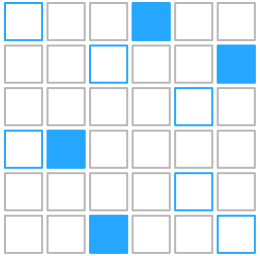


Recsys

Train



Test



Which metric is the best to evaluate the model performance?

Decision-support metrics

* MSE
* AUC
* Log Loss
* HR@k
* **MAP@k**
* nDGC@
* ,...

Recommendation centric

* Coverage
* **Diversity** ("filter bubble" problem)
* Confidence

User centric

* Trustworthiness
* Novelty
* Serendipity

System centric

* Robustness
* Scalability
* Stability
* Privacy
* Latency
* ...

Business metrics

* CTR
* CVR
* AOV
* Engagement
* Frequency
* Retention
* ...

Online vs Offline Evaluation

In the RS community there are a general agreement that **evaluate a RS it's close to impossible without having a live system** to test it (AB test).

Still it's important to know if your recommender system is going in the write direction.

Table 1: Offline & online metrics of different models. Online Acquisition Gain is relative to the control.

Model	Offline AUC	Online Acquisition Gain
Wide (control)	0.726	0%
Deep	0.722	+2.9%
Wide & Deep	0.728	+3.9%

WDL - Google Apps store - app recommendations

Table 4: Offline AUCs and online CTR gains of different methods. Online CTR gain is relative to the control group.

Methods	Offline AUC	Online CTR Gain	Average RT(ms)
WDL	0.7734	-	13
WDL(+Seq)	0.7846	+3.03%	14
DIN	0.7866	+4.55%	16
BST(b = 1)	**0.7894**	**+7.57%**	20
BST(b = 2)	0.7885	-	-
BST(b = 3)	0.7823	-	-

BST - Alibaba CTR prediction

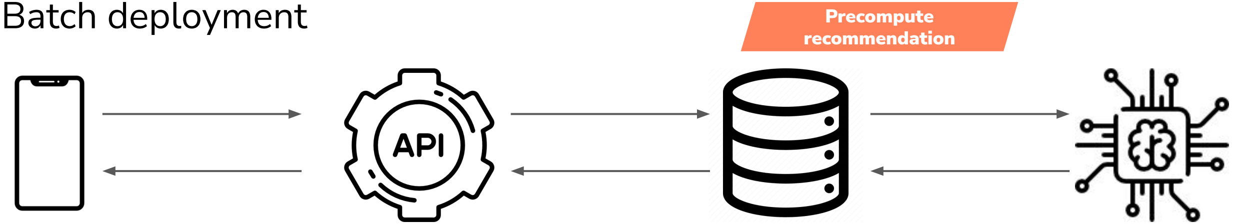
Is the feature like the past?

What happen if the marketing team make a special promotion and a lot of people buy some products? Are this signal useful for the next recommendation?

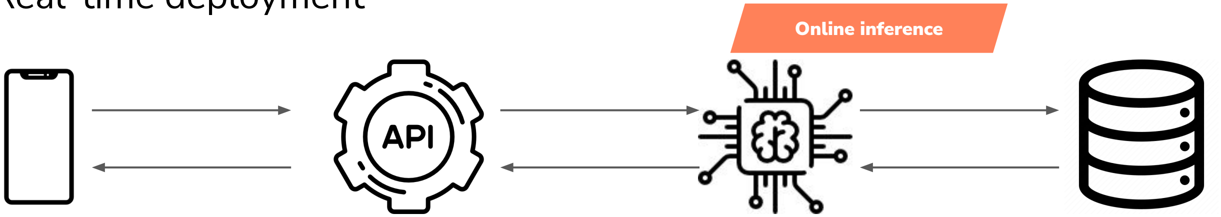
RS to production

As any other ML type, RecSys are a software application that live in production

Batch deployment

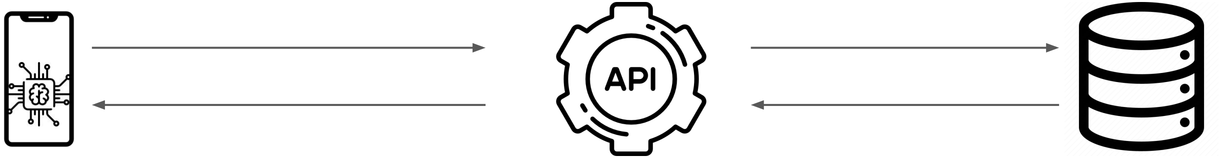


Real-time deployment



Retrieve candidate from thousand or millions of item is extremely computationally intensive

Edge deployment



Recommendation system not just recommendation model

Recommendation system is much more complex that recommendation algorithms/models

Two-stage Recommender Systems

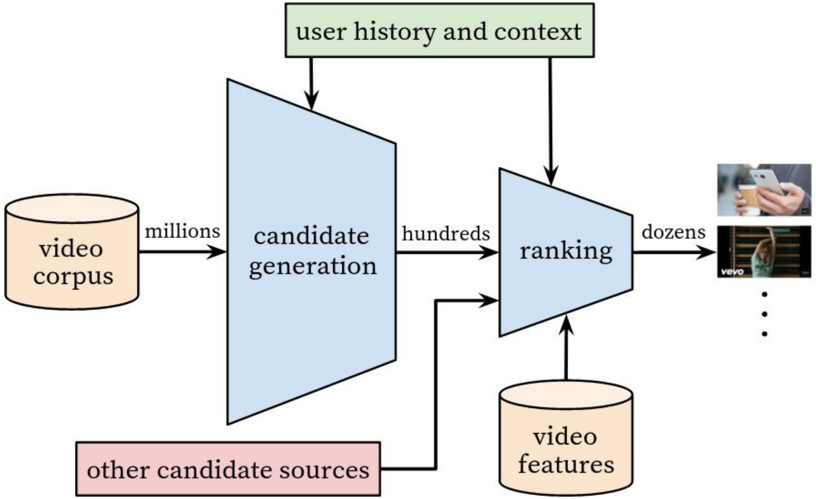
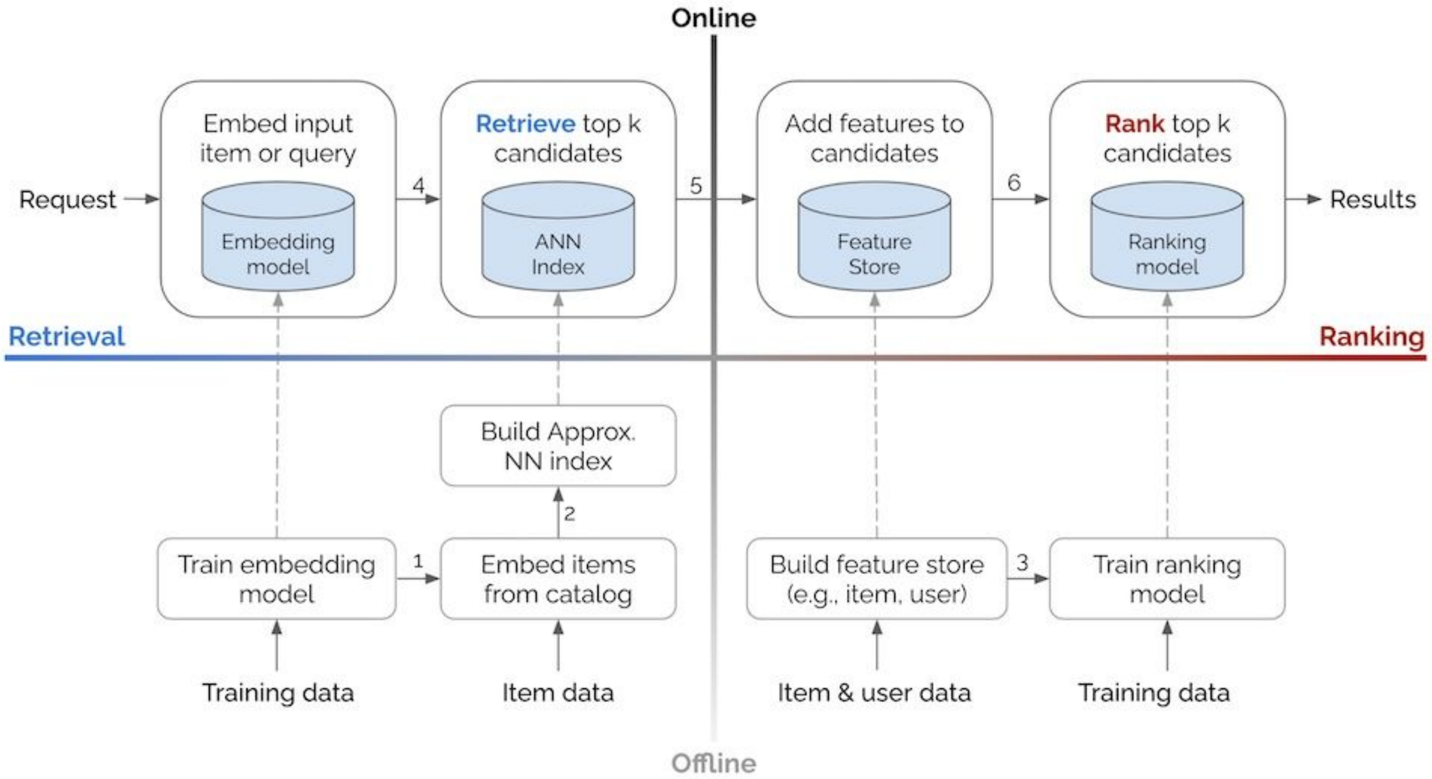


Figure 2: Recommendation system architecture demonstrating the "funnel" where candidate videos are retrieved and ranked before presenting only a few to the user.

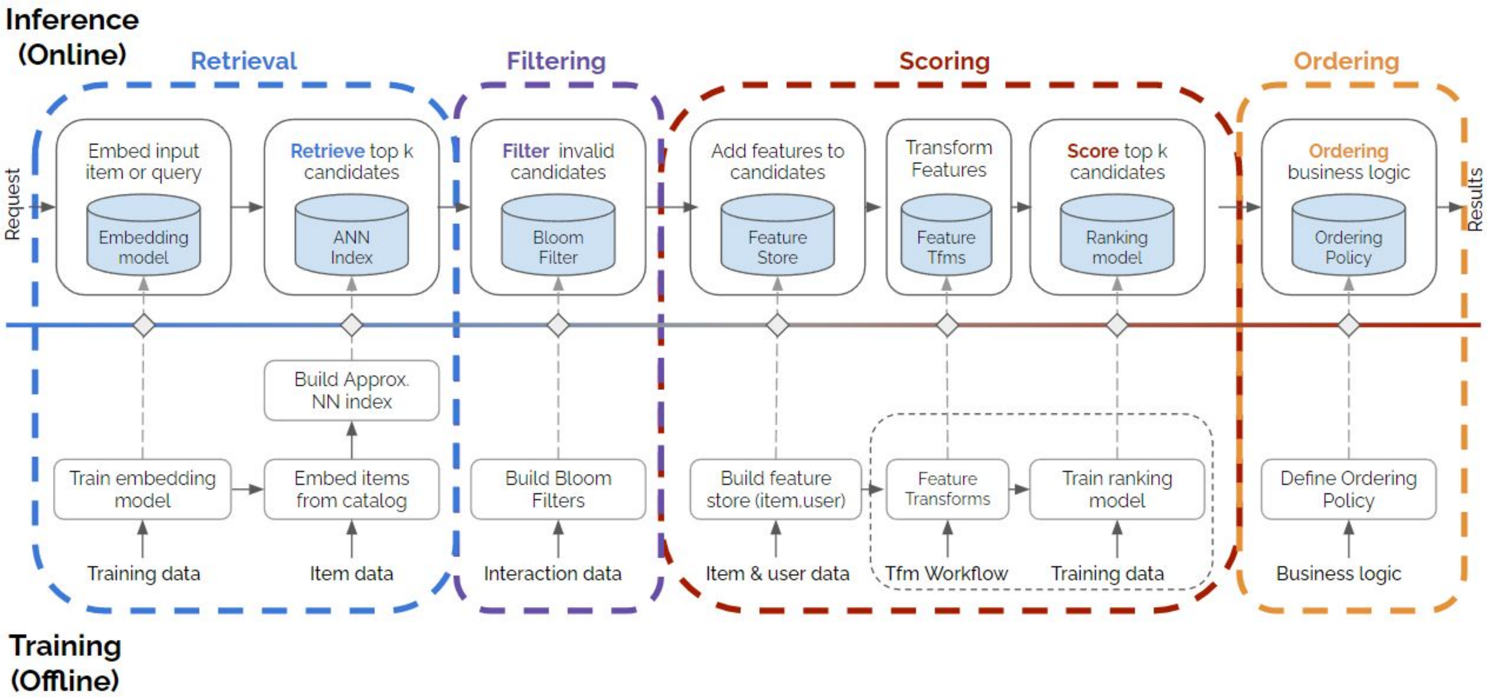
[Deep Neural Networks for YouTube Recommendations]

Two-stage Recommender Systems



[System Design for Recommendations and Search (Eugene Yan)]

Four-stage Recommender Systems



[Moving Beyond Recommender Models - Even Oldridge (NVIDIA)]

Productionize

End2End flow

Tricky to train

Large scale model with **high-cardinality sparse features**

Tricky to evaluate

Offline metrics can be **highly misleading**

Tricky to deploy

Efficiently retrieval for acceptable latency

Productionize

End2End Frameworks

TF Recommenders

Merlin

TorchRec

“There is more art than science in the development of Recommendation Systems …”

Deep Neural Networks for YouTube Recommendations

Productionize

Our Product recommendations strategy

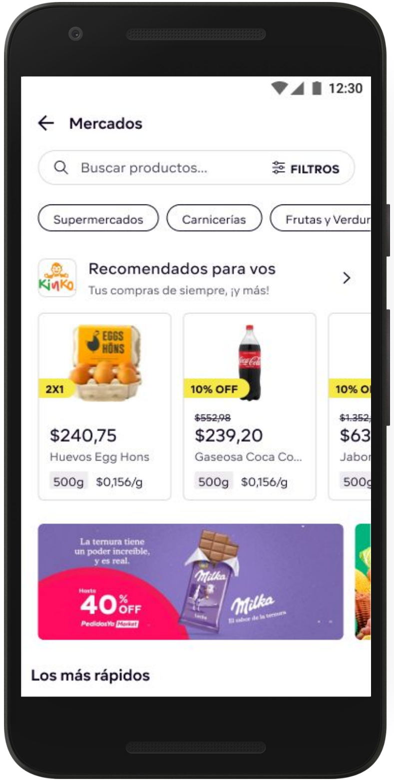
user_id

item_id

order_id

TensorFlow Recommenders

NVIDIA
TRITON INFERENCE SERVER



Demo

Frameworks & tools

Name	Link
Surprise	https://github.com/NicolasHug/Surprise
LightFM	https://github.com/lyst/lightfm
Implicit	https://github.com/benfred/implicit
Spotlight	https://github.com/maciejkula/spotlight
TF Recommenders	https://github.com/tensorflow/recommenders
TensorRec	https://github.com/jfkirk/tensorrec
RecBole	https://github.com/RUCAIBox/RecBole
Collie Recs	https://github.com/ShopRunner/collie_recs
DeepCTR	https://github.com/shenweichen/DeepCTR
Nvidea Merlin	https://github.com/NVIDIA-Merlin/Merlin

Name	Link
RecPack	https://gitlab.com/recpack-maintainers/recpack
RecList	https://github.com/jacopotagliabue/reclist