

## Final presentation

### Subject 1 - Product recommendation

*March 22<sup>nd</sup>, 2021*

Eloi ALARDET - Joël GARDE - Mathieu JUTTET

# Agenda



INSTITUT  
POLYTECHNIQUE  
DE PARIS

1. Exploring the dataset
2. Building the model
3. Obtained results

# 1. Exploring the dataset



Exploring the data set:

- 12 months of data - 2019
- 116k clients
- 12k products
- 140M rows of data in total (120M for training and 20M for testing) - 3 Go

# 1. Exploring the dataset



Exploring the data set:

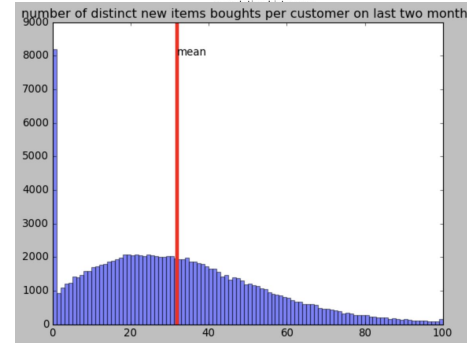
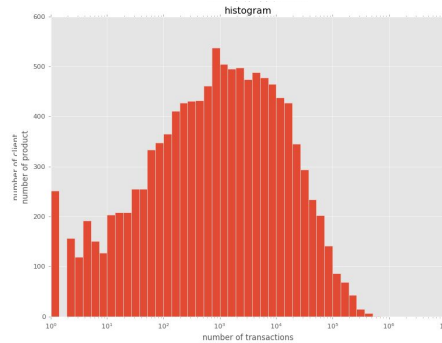
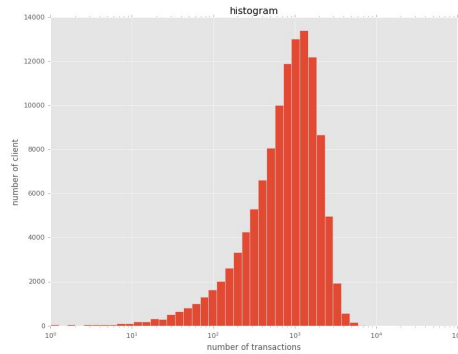
- 12 months of data - 2019
- 116k clients
- 12k products
- 140M rows of data in total (120M for training and 20M for testing) - 3 Go

**Reasonably big**

# 1. Exploring the dataset

## Clients:

- 1 000 different products bought per year in average
- In average +20% of new products every 2 months

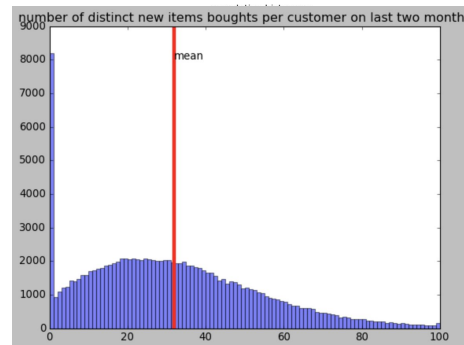
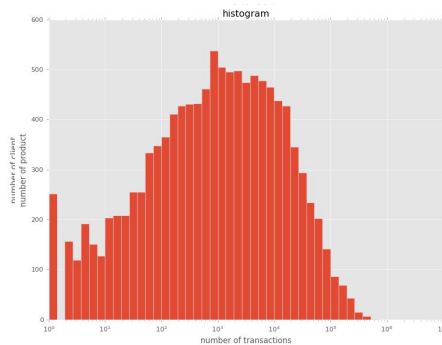
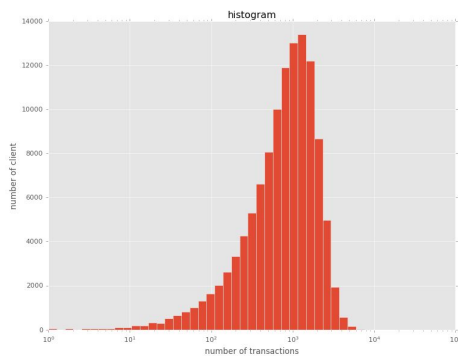


# 1. Exploring the dataset

## Clients:

- 1 000 different products bought per year in average
- In average +20% of new products every 2 months

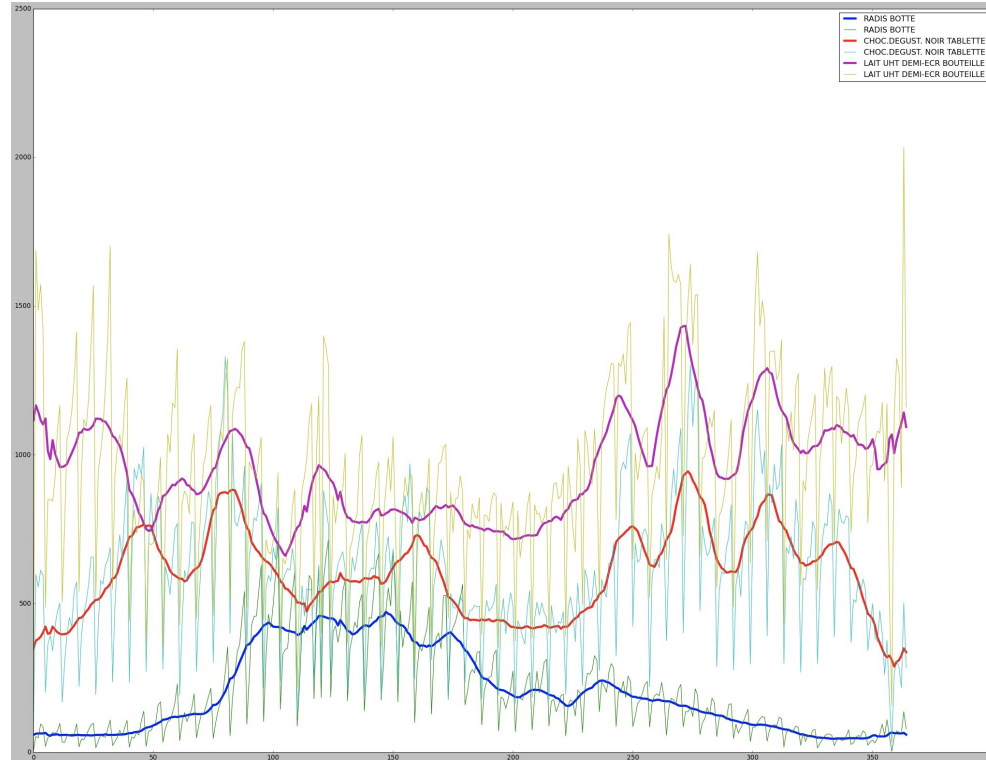
**Prediction difficult to make**  
**Need for recognising similarities**



# 1. Exploring the dataset

Products:

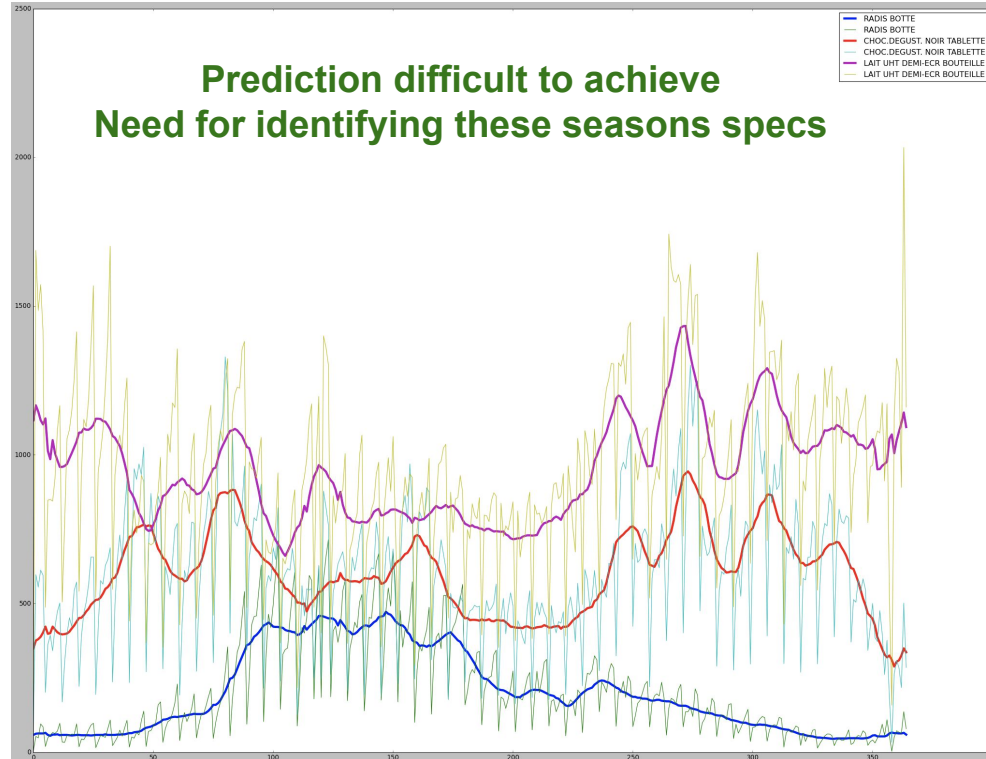
- Seasonality due to discount periods, seasons and main periods of the year



# 1. Exploring the dataset

Products:

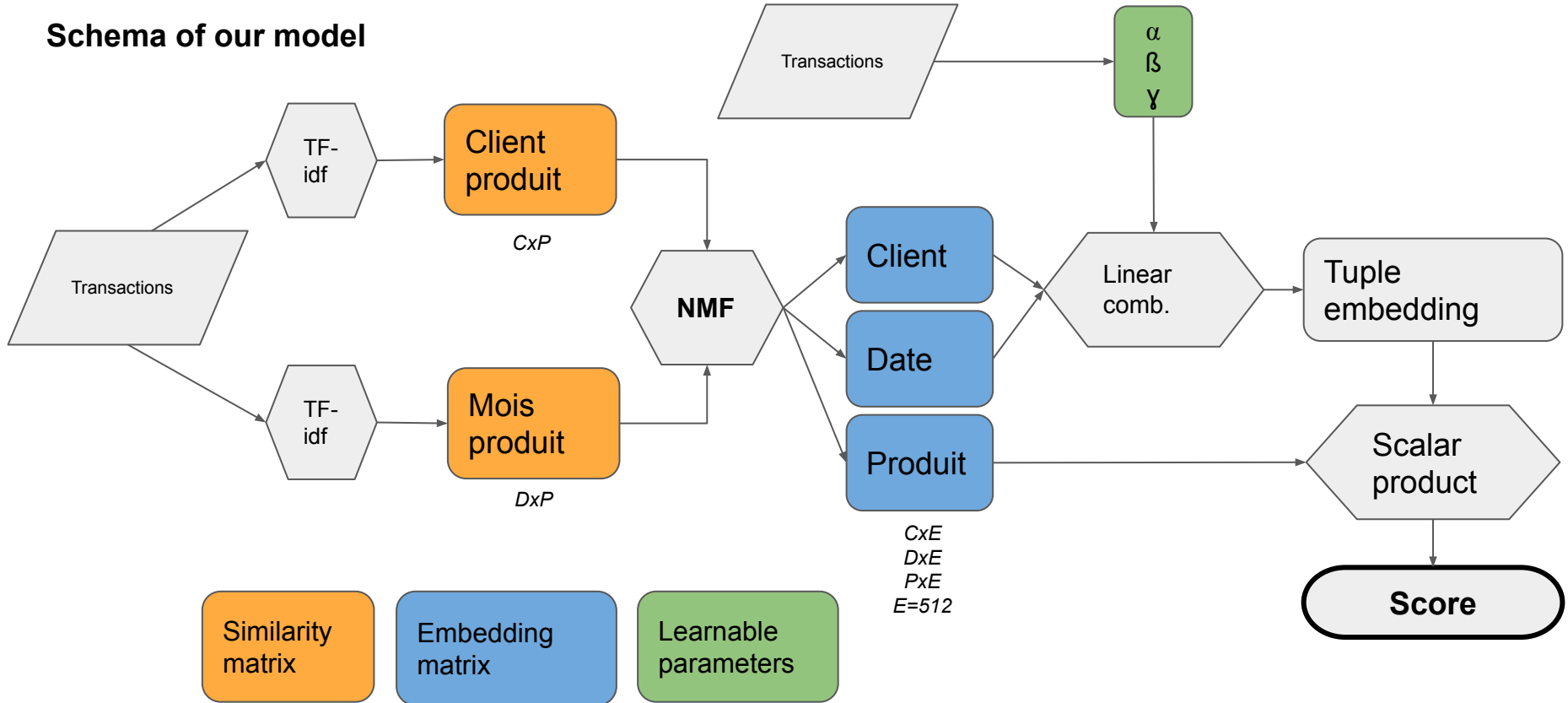
- Seasonality due to discount periods, seasons and main periods of the year





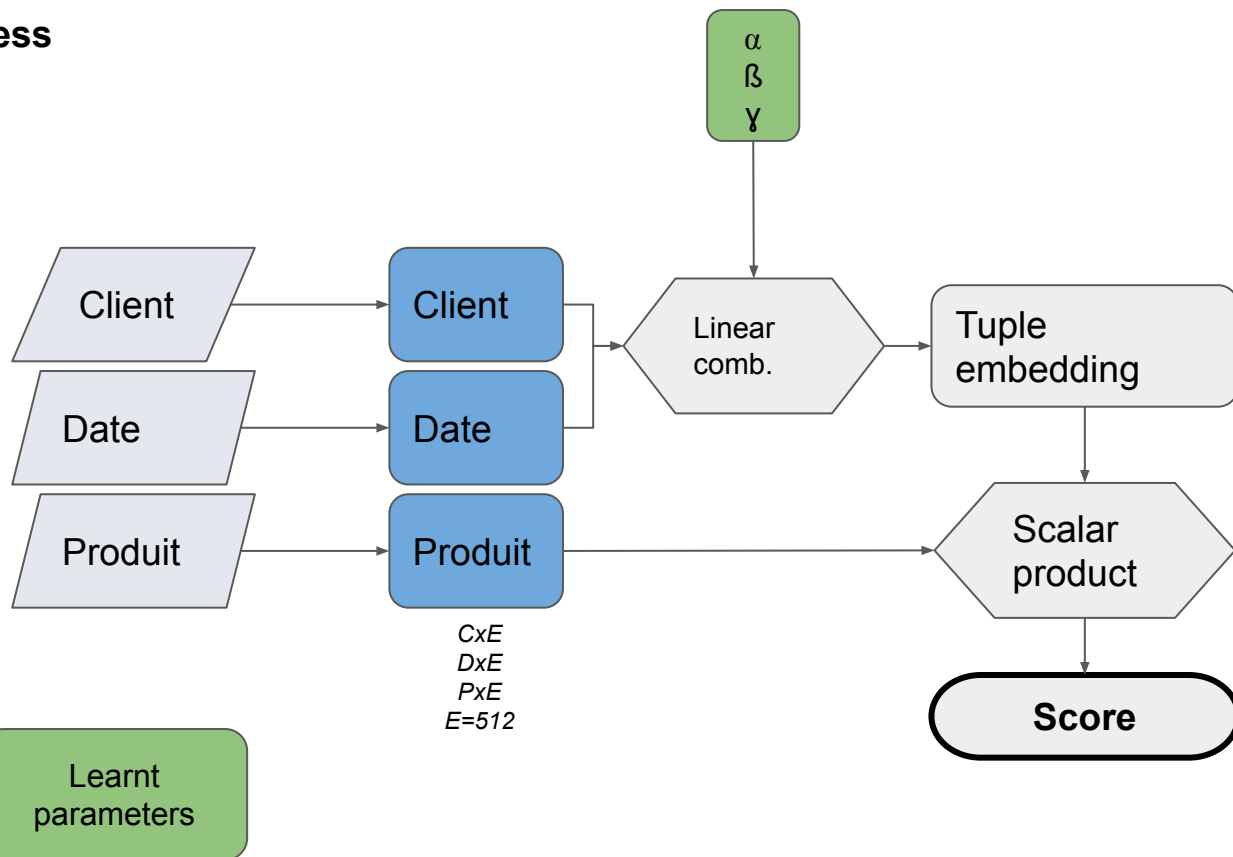
## 2. Building the model

### Schema of our model



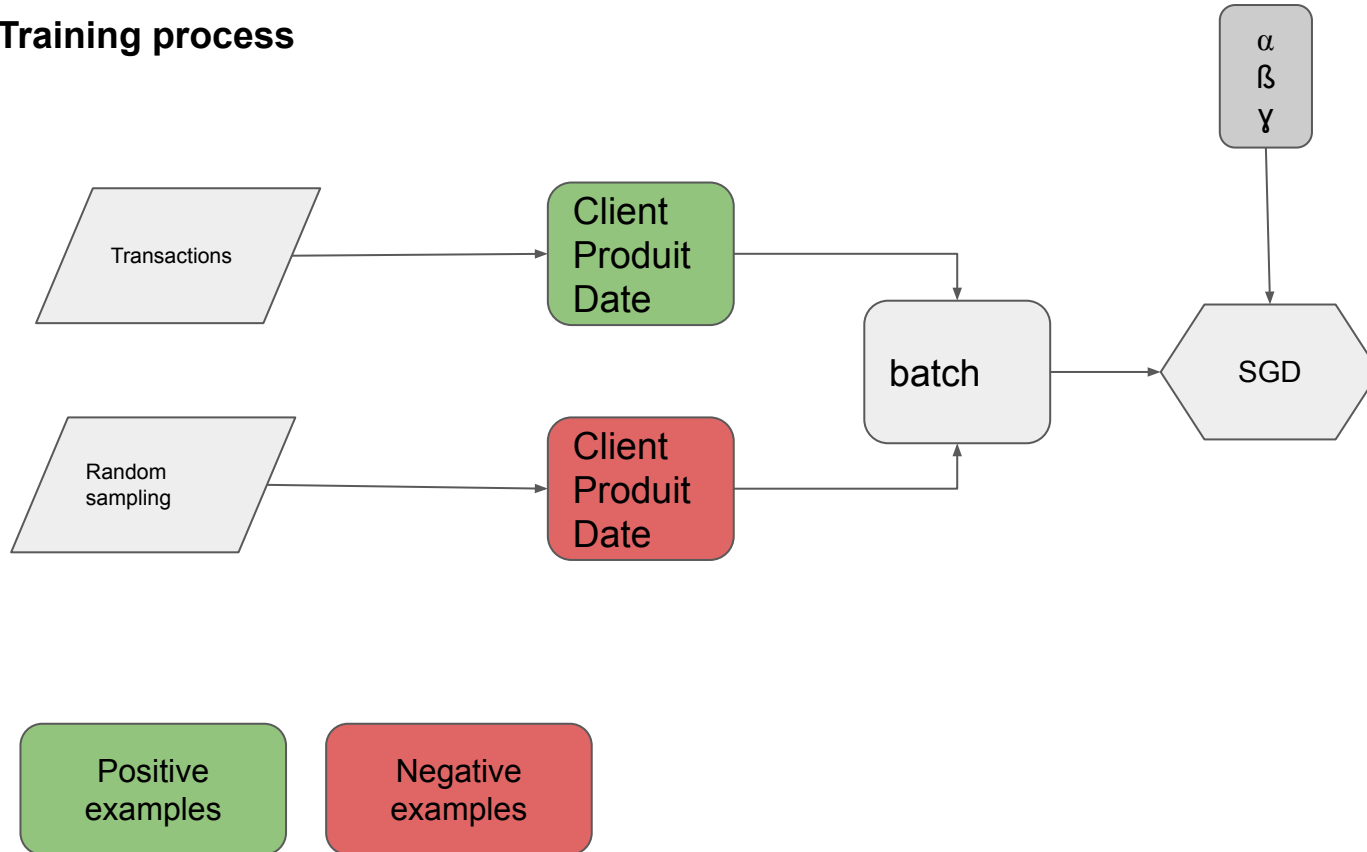
## 2. Building the model

### Prediction process



## 2. Building the model

### Training process



## 2. Building the model



### Motivation of using matrix decomposition

- ✓ Using two embeddings gives the user control over the features
- ✓ The model can predict new item to be bought
- ✓ Handles large amounts of data
- ✓ Can be used in a pipeline for other tasks

### 3. Obtained results

Predictions \ Truth	True	False
True	819 023 = TP	1 413 056 = FP
False	1 587 644 = FN	15 387 875 = TN

Model	Precision	Recall	F1-score
Trained model	<b>0.367</b>	<b>0.340</b>	<b>0.353</b>
“Dumb” model	0.125	1.00	0.222

Computation time : 2 minutes (!)

# Conclusion



- Possible improvements: the way we implement similarity matrices can evolve to integrate more features (importance of promotions, current trends of shopping...)
- Overall, a really interesting project, representative of what we can encounter in companies: GCP environment, problematics.
- We think our solution is innovative in its way and scalable
- We would have loved more time to refine our solution.