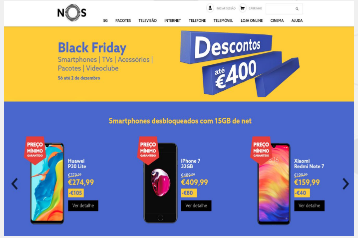
Providing relevant recommendations beyond the explored frontiers

Kelwin Fernandes, Ph.D.
CEO @ NILG.AI
20191203

in partnership with

NEXT BEST OFFER @ NOS (NBO)

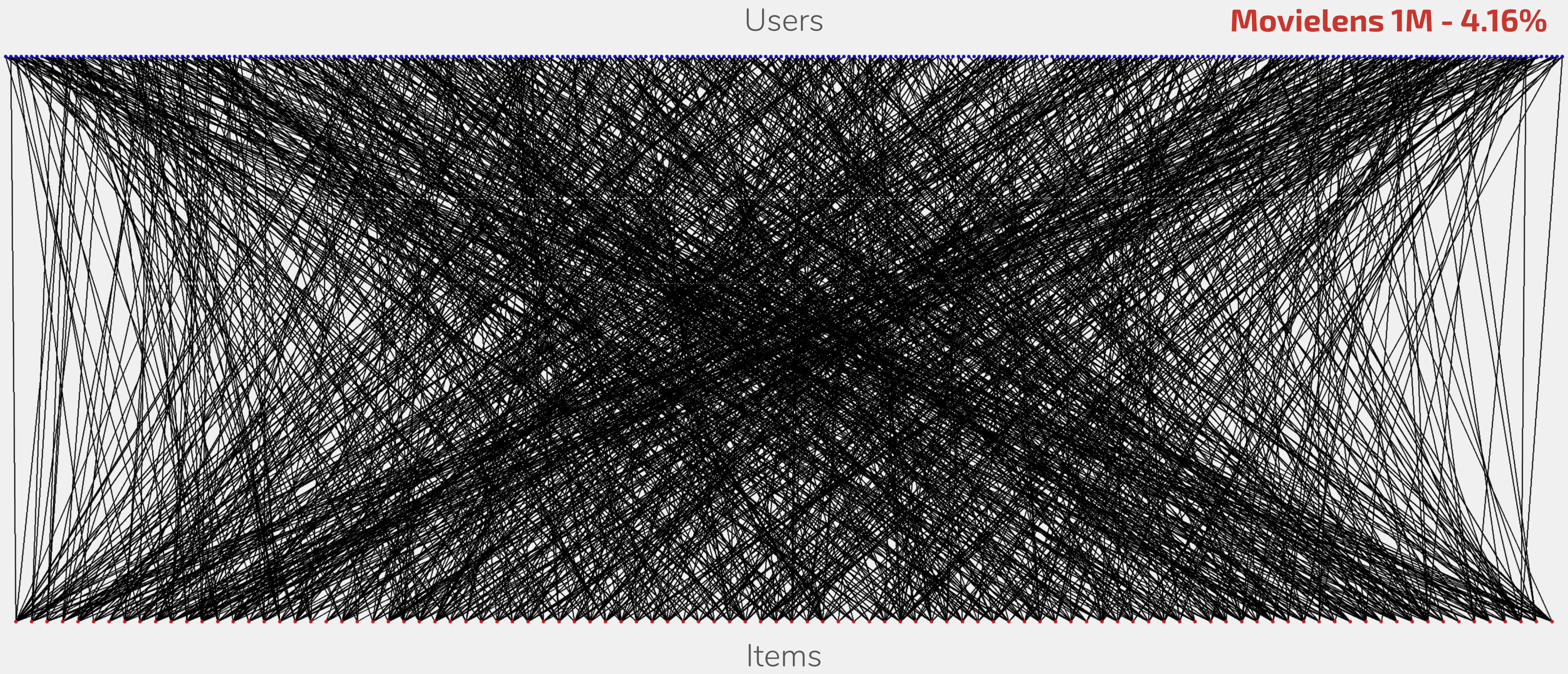
- **Not-so-**digital interactions
  - Channels:
    - Visits to the stores
    - Outbound/Inbound calls
    - Door-to-door sales
    - Website
  - Interactions are SUPER **scarce & noisy**
    - 0-3 calls (0-1 sales) per client/year



- Tons of **product combinations**
  - TV + NET + Phone + SIM + Data + others
  - ~15k potential combinations

NBO @ NOS - Cold Start (winter is coming!)

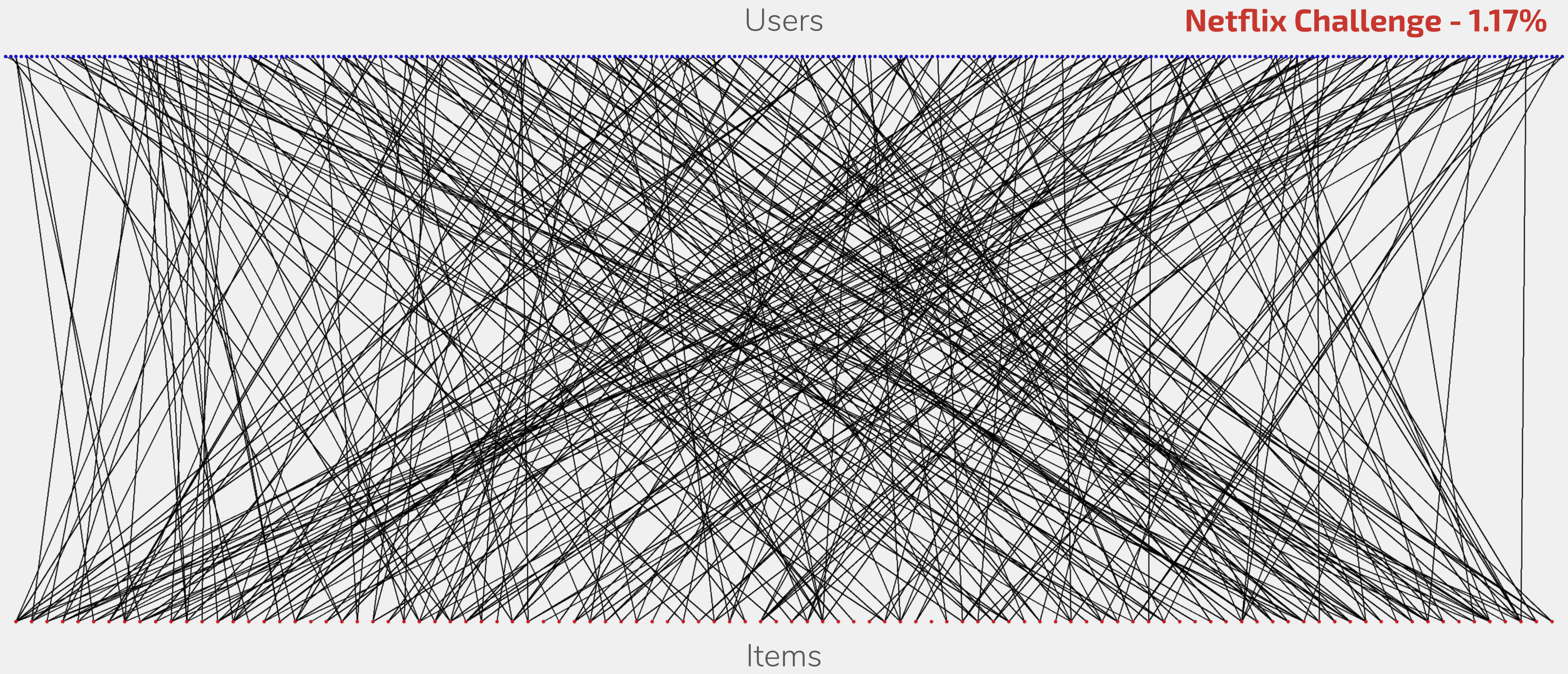
in partnership with NOS



NBO @ NOS - Cold Start (winter is coming!)

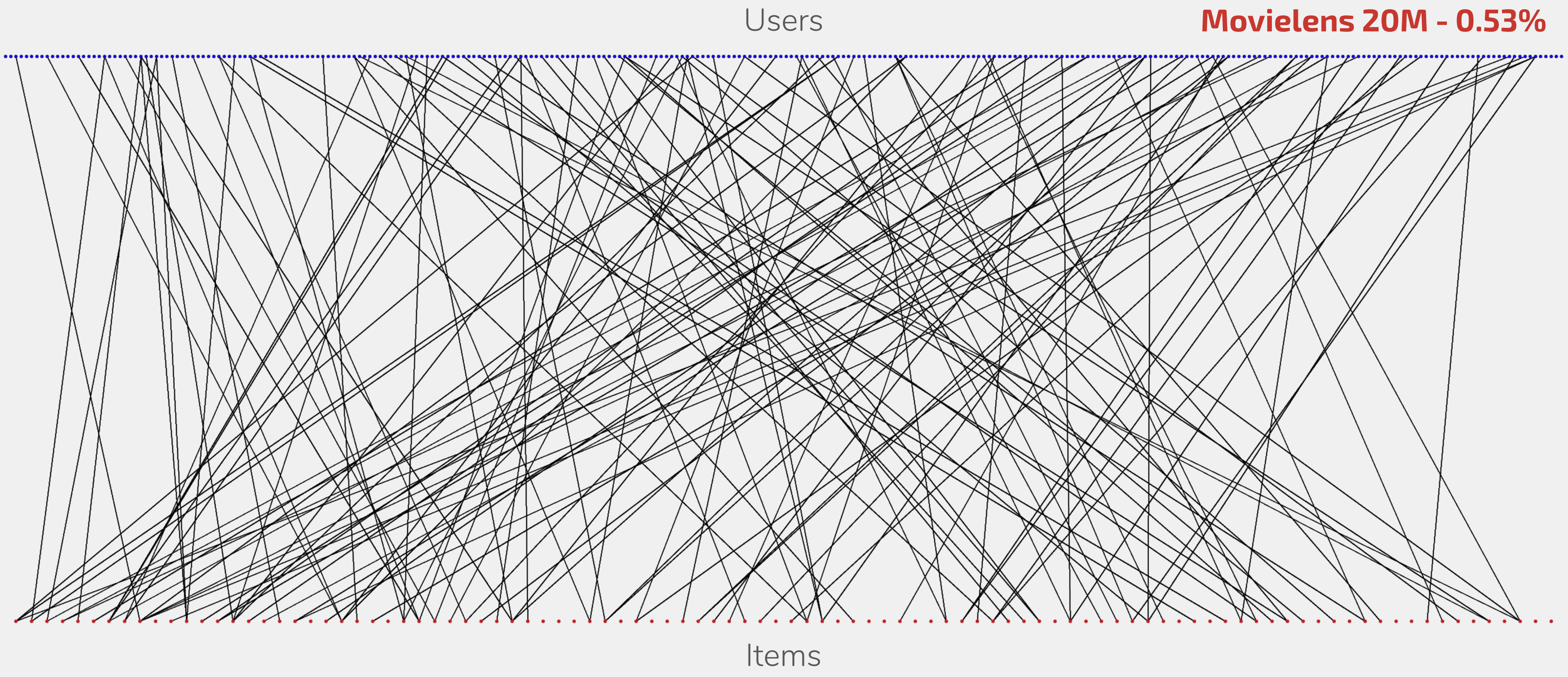
in partnership with NOS

Netflix Challenge - 1.17%



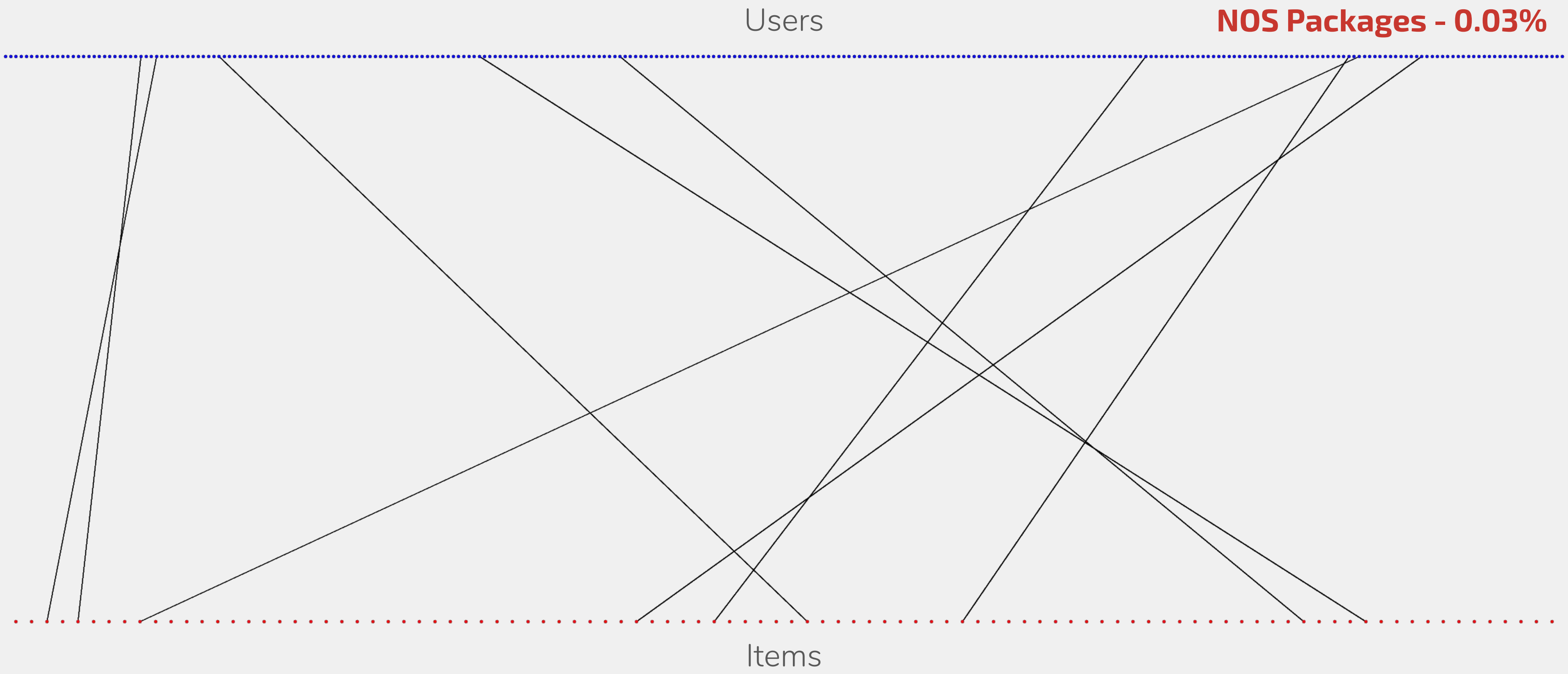
NBO @ NOS - Cold Start (winter is coming!)

Movielens 20M - 0.53%



NBO @ NOS - Cold Start (winter is coming!)

nilg.ai in partnership with NOS

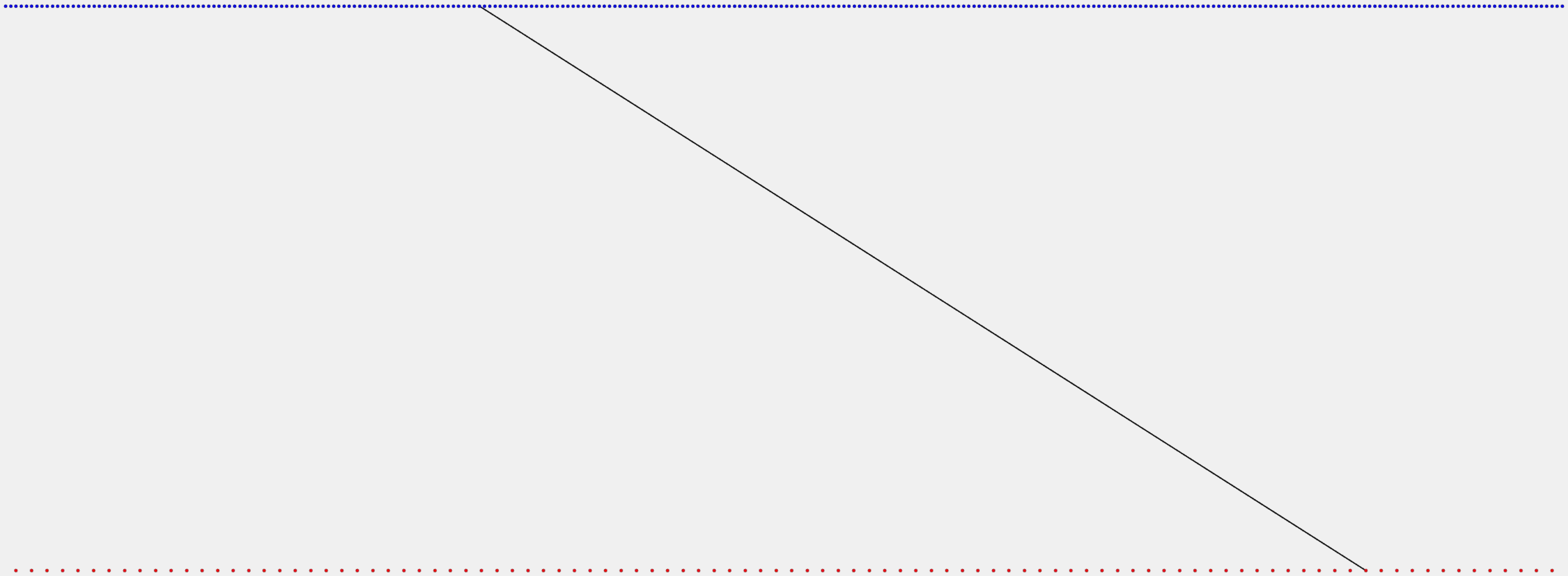


in partnership with NOS

NBO @ NOS - Cold Start (and forget about getting any hotter)

Users

NOS Packages (Sales) - 0.0075%



Items

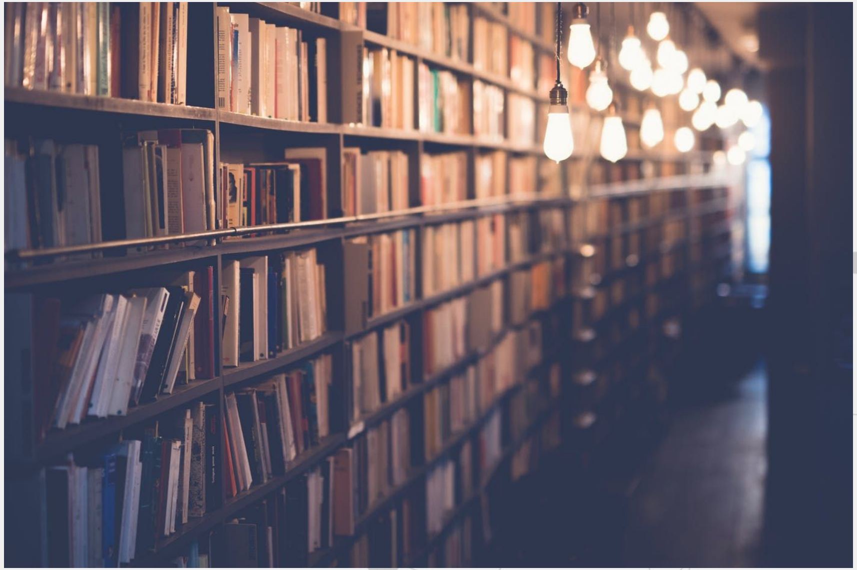
NBO @ NOS - How to handle it?

- Multiple approaches from the literature:
  - Basket Analysis
  - Collaborative Filtering
    - User-User, Item-Item, …
    - Matrix Factorization

- Long story short:
  - Yes, we tried them!
  - No, they didn't work!

NBO @ NOS - The initial approach

- ## Introducing Domain Knowledge
  - You know your customer behavior
  - You can describe your products



- ## Content-Based Models
  - Recommendation as **binary classification**
    - `f :: user x offer → success`
  - Using features such as:
    - Customer behavior
    - Product categorization
    - and of course... previous interactions with the customer

NBO @ NOS - The initial approach

- **Introducing Domain Knowledge**
  - You know your customer behavior
  - You can describe your products

- **Content-Based Models**
  - Recommendation as **binary classification**
    - `f :: user x offer → success`
  - Using features such as:
    - Customer behavior
    - Product categorization
    - and of course… previous interactions with the customer

1st advice

Cold Start + Domain Knowledge ⇒ Content-based RecSys

in partnership with NOS

NBO @ NOS - The initial approach

- **Project Manager:** I want a model to recommend a new product
  - e.g.,a new box model, NOS Kids, App NOS, cellphone with steroids, etc
- **Data Scientist:** Here is my shopping list
  - Some **training data**
    - Build pilots on call-centers:
      - i.e., stop what you're doing, build a script, train agents
    - Start bothering thousands of clients to get data
  - Three-four months to **train/evaluate a new model**
  - Lots of **discussions** to understand if the model is ready to go
  - + some more time to include this model in the **decision pipeline**

nih.ai in partnership with NOS

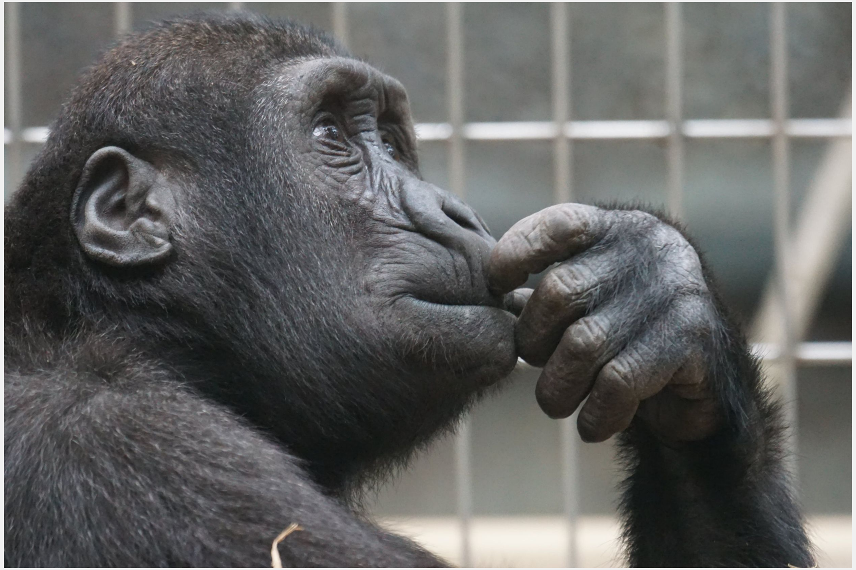
NBO @ NOS - The initial approach

- One model per offer type x customer segment
  - Months of development and 11 models later...

4P @ 3P | + Cartões | + Dados | + Integrados

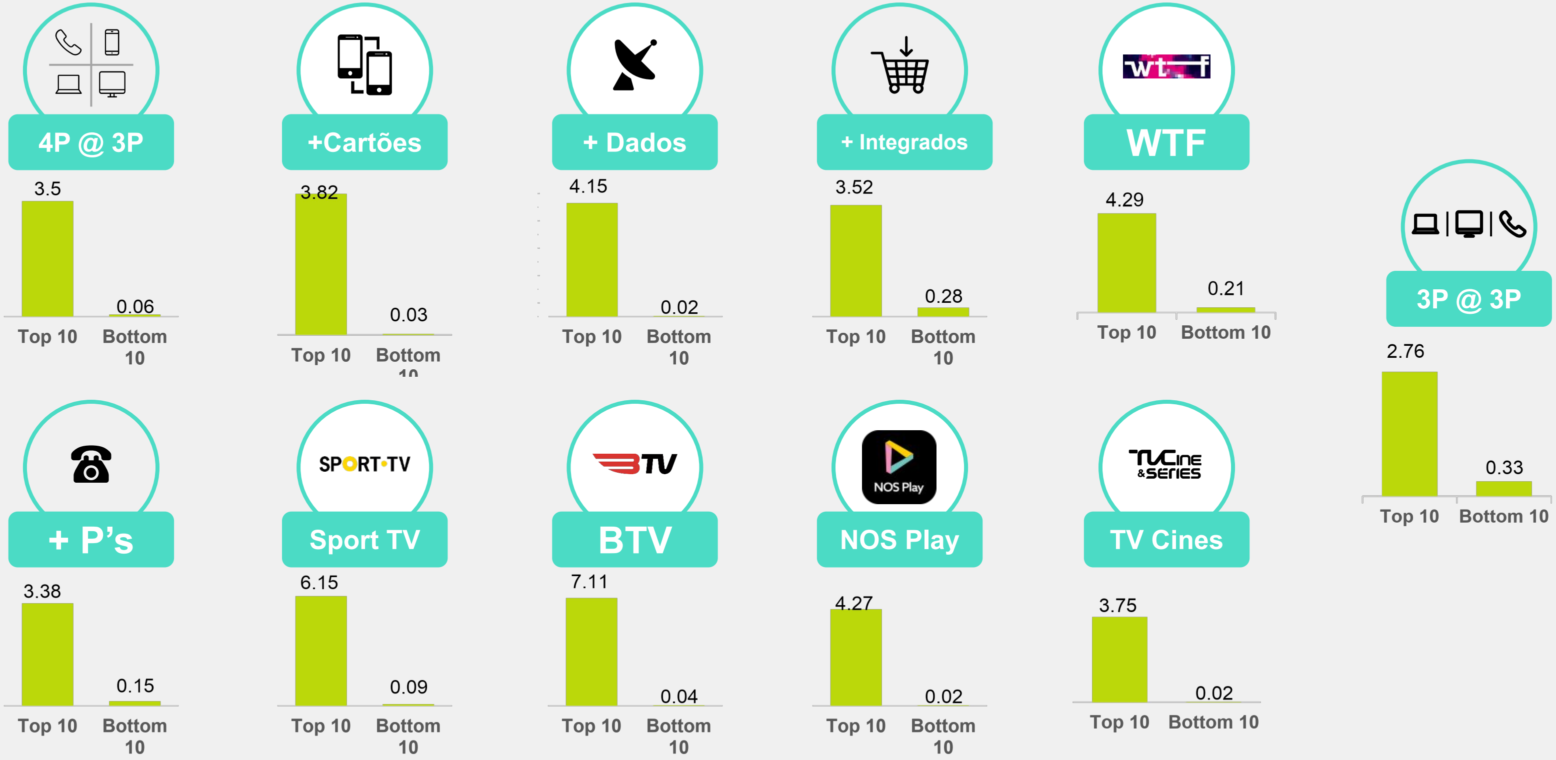
+ P's | Sport TV | BTV | NOS Play

WTF | TV Cine | 3P @ 3P



A zoo of thinking animals trapped in cages. That you needed to feed, to pet and to keep alive...

NBO @ NOS - The initial approach



NBO @ NOS - Global Propensity Model

nihai

in partnership with NOS

- Instead of building a new model per requirement:

  $f_{\text{product-specific}} :: \text{user} \rightarrow \text{sales success}$

- Build a **global model** to map all your sales interactions:

  $f_{\text{global}} :: \text{user} \times \text{origin} \times \text{offer} \rightarrow \text{sales success}$

We had to put them down!

NBO @ NOS - Global Propensity Model

NBO @ NOS - Global Propensity Model

- Demographics
- Product-usage patterns
- Relevance of each product on the user-usage
  - e.g., top 10% data-usage vs 30% TV
- Previous Interactions (service requests, tickets, etc)
- etc

NBO @ NOS - Global Propensity Model

- Characterization of the current subscription:
  - Price
  - Loyalty period
  - Number of TV channels
  - Number of SIM cards
  - Data plafond
  - NET speed
  - Box type
  - etc

NBO @ NOS - Global Propensity Model

- Characterization of the new offer:
  - Price
  - Loyalty period
  - Number of TV channels
  - Number of SIM cards
  - Data plafond
  - NET speed
  - Box type
  - etc

in partnership with NOS

NBO @ NOS - Global Propensity Model

- Tons of feature-engineering:
  - Pairwise Interactions:
    - Absolute Deltas: Offer feature - Current feature
      - +2€, +1GB
    - Relative Deltas: Offer feature / Current feature
      - +10% price, +50% plafond
    - User consumption vs. Offer delta
  - Ranking, Trend analysis, etc

NBO @ NOS - Global Propensity Model

- Tons of feature-engineering:
  - Pairwise Interactions:
    - Absolute Deltas: Offer feature - Current feature
      - +2€, +1GB
    - Relative Deltas: Offer feature / Current feature
      - +10% price, +50% plafond
    - User consumption vs. Offer delta
  - Ranking, Trend analysis, etc

2nd advice

Sometimes dividing doesn't mean conquering

in partnership with

NBO @ Any Subscription-based Company

Characterize car leasing:

- Price
- Brand
- Power
- Fuel type
- Efficiency
- Seats
- etc

Characterize streaming:

- Price
- Number of devices
- Access to content
- Monthly/Yearly
- etc

Characterize software:

- Price
- Number of users
- Add-ons
- Memory limits
- etc

NBO @ NOS - Dataset

- **Target task:** relevant outbound interactions
  - Cleanest data we have: sales vs. non-sales
  - Still negatives are somehow uncertain:
    - Haven't decided yet
    - The agent didn't reach to present the offer

- Including additional data:
  - Organic sales:
    - Only positive values
    - Proxy task: sales vs. fake opponent
  - Clients out-of-loyalty period

Instance-based transfer learning

NBO @ NOS - Dataset

- **Target task:** relevant outbound interactions
  - Cleanest data we have: sales vs. non-sales
  - Still negatives are somehow uncertain:
    - Haven't decided yet
    - The agent didn't reach to present the offer

3rd advice

Merge alternative data sources related to your target task

(transfer learning isn't just weight transfer + fine-tuning)

- Including additional data:
  - Organic sales:
    - Only positive values
    - Proxy task: sales vs. fake opponent
  - Clients out-of-loyalty period

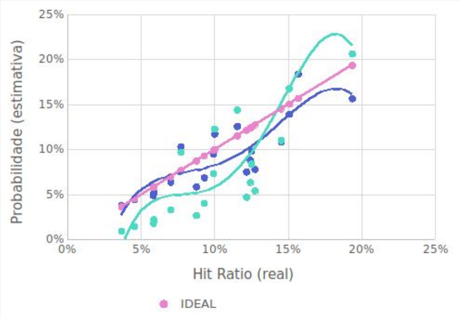
Instance-based transfer learning

NBO @ NOS - Decision Function

- Let's O={o_0, o_0, ..., o_n} be a set of offers with price PVP_0, PVP_1, ..., PVP_n respectively.
- Let's P(C, o) be the probability of customer C buying offer o.

- Maximize Average Revenue Per User (ARPU)

$$\underset{o_i \in O}{\operatorname{argmax}} P(C, o_i) \times PVP(o_i)$$



Probability Calibration:

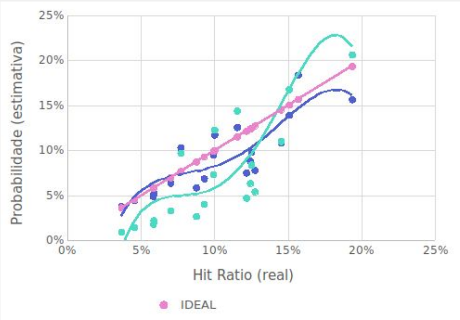
Isotonic Regression, Logistic Regression, ...

NBO @ NOS - Decision Function

- Let's $O=\{o_0, o_0, \ldots, o_n\}$ be a set of offers with price $PVP_0, PVP_1, \ldots, PVP_n$ respectively.
- Let's $P(C, o)$ be the probability of customer C buying offer o.

- Maximize Average Revenue Per User (ARPU)

$$\underset{o_i \in O}{\operatorname{argmax}}\ P(C, o_i) \times PVP(o_i)$$



4th advice

Calibrate your probabilities in any decision-making process, especially if you manipulated the distributions!

Probability Calibration:

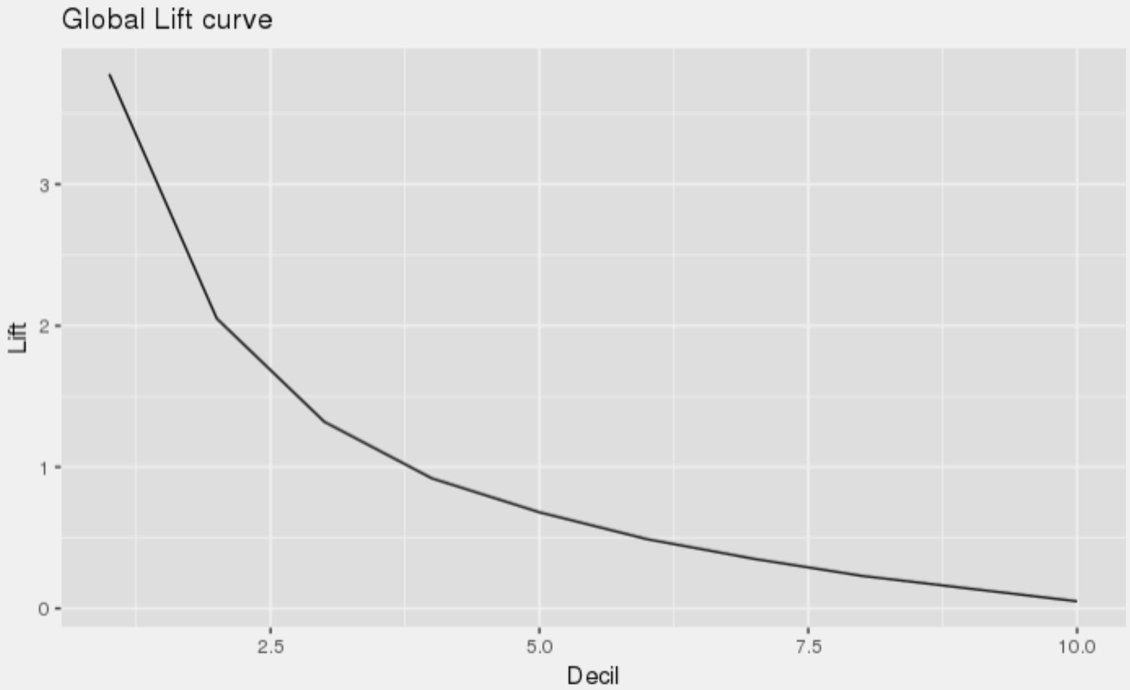
Isotonic Regression, Logistic Regression, ...

Feature Importance by category

Feature Importance by category		
Index	Feature	Gain
1	FeatEng: Relative deltas of origin vs. Offer components	55%
2	Price-related features	51%
3	FeatEng: Pairwise Interactions	50%
4	Behavioral data	38%
5	FeatEng: Absolute deltas of origin vs. Offer components	7%
6	Previous interactions	2%

in partnership with

Cross-Validation Performance



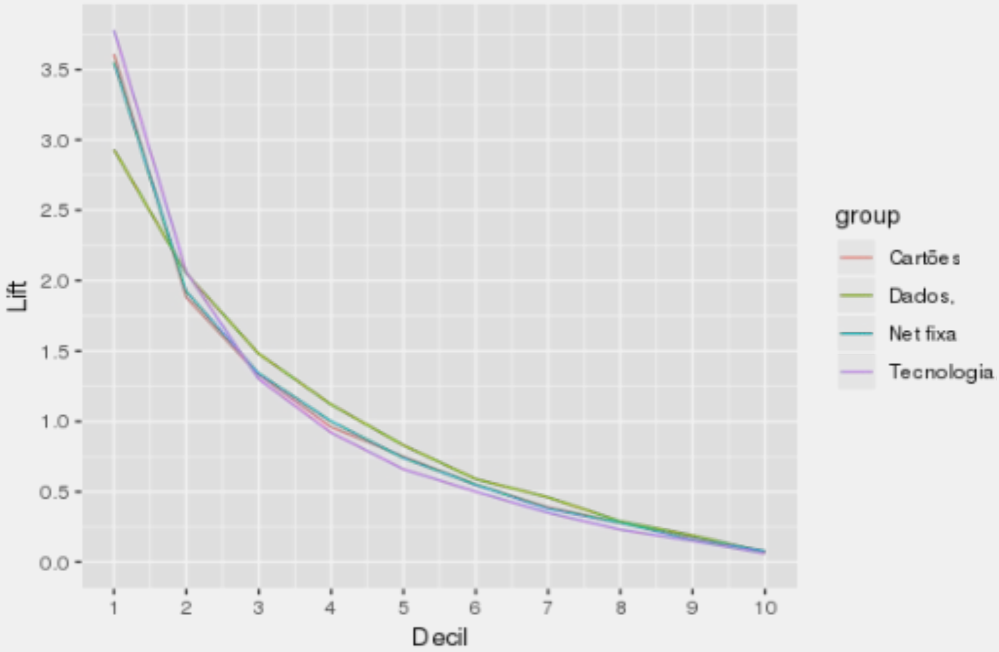
3.77

Lift top 10% vs base

85.3%

ROC AUC

Cross-Validation Performance



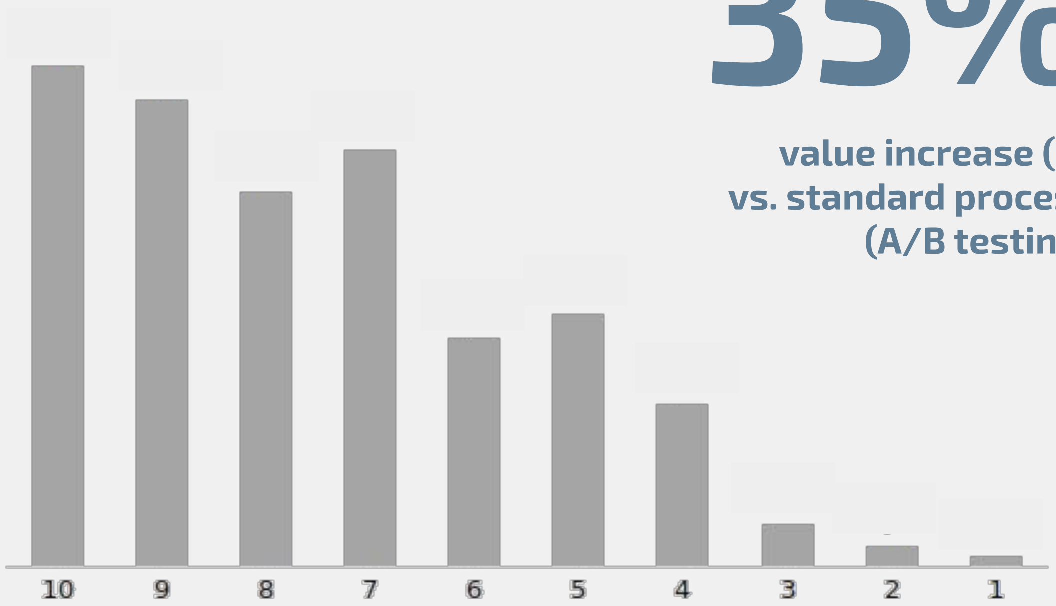
3.77

Lift top 10% vs base

85.3%

ROC AUC

Real-life results

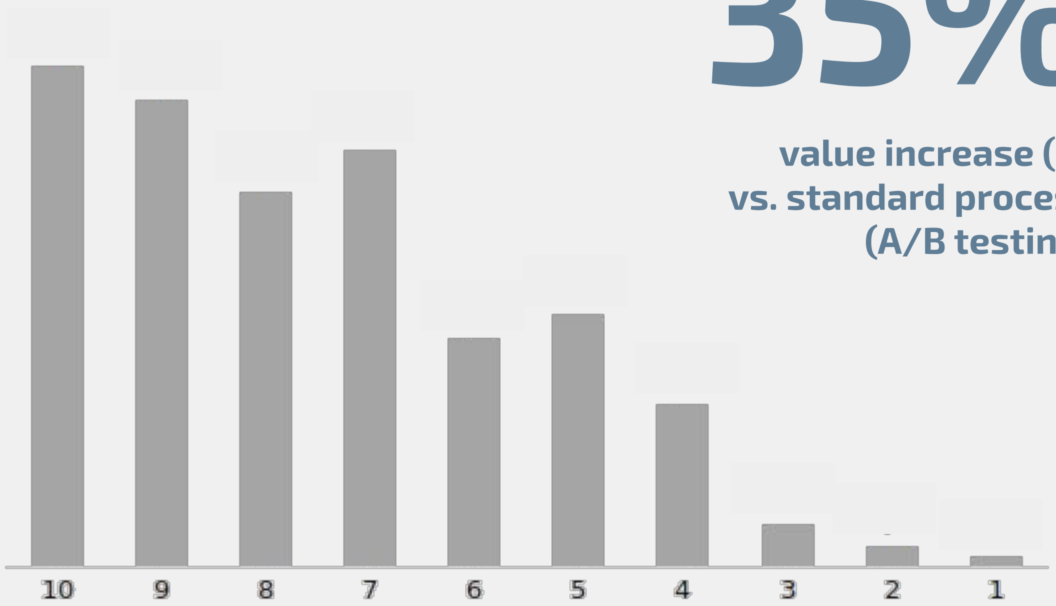


35%

**value increase (€)
vs. standard process
(A/B testing)**

nihai.ai in partnership with NOS

Real-life results



35%

value increase (€)
vs. standard process
(A/B testing)

Go back to the Caribbean, you can retire…

wait, not so fast!

nilg.ai in partnership with NOS

Exploring beyond the known frontiers

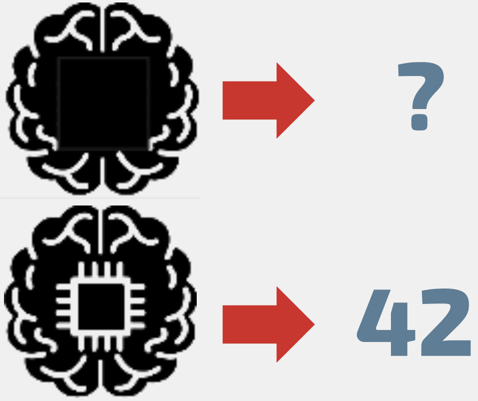
- You said your model would predict **every type of sale**...
- I want to predict on some new products I haven't tested before...

Training

2 x 2 = 4
4 x 3 = 12
5 x 4 = 20
8 x 7 = 56
...

Test

真 久



Exploring beyond the known frontiers

- **Uncertainty Estimation**
- 3 types of uncertainty:
  - **Aleatoric**

Randomness in the process:

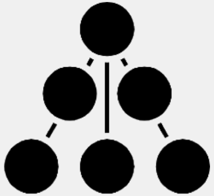
- Noise
- Unobserved variables
- Context

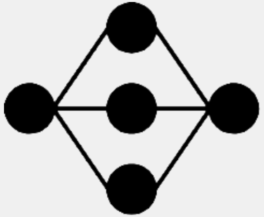
Prof	PVP	PF	Data	Cards	Airtime	Aud. Football	Data usage	Delta Cards	Delta PVP	Kids	Debts
3P	31	23	1GB	0	300	10k	80%	1	15	?	?
4P	40	12	5GB	2	600	0k	60%	1	10	?	?
4P	50	6	10GB	1	1000	500	60%	2	7	?	?
3P	29.9	20	0GB	0	500	0k	90%	0	5	?	?

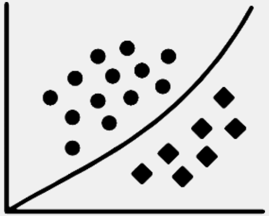
This is why we don't have 100% accuracy even with a perfect model.
We are used to measure this!

Exploring beyond the known frontiers

- **Uncertainty Estimation**
- 3 types of uncertainty:
  - Aleatoric
  - **Epistemic**

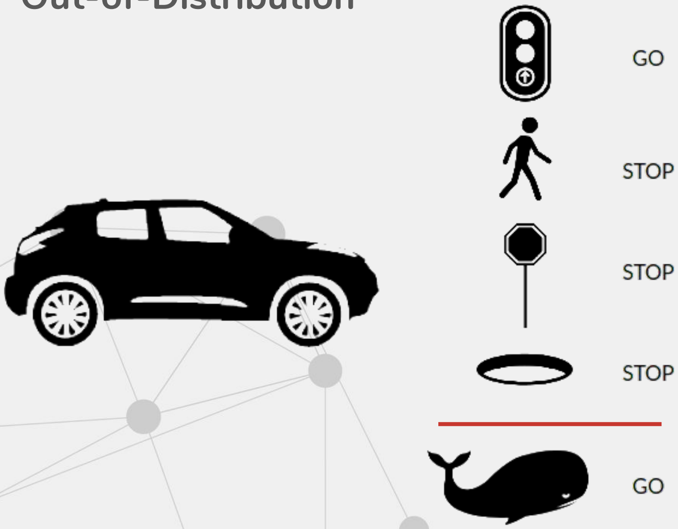


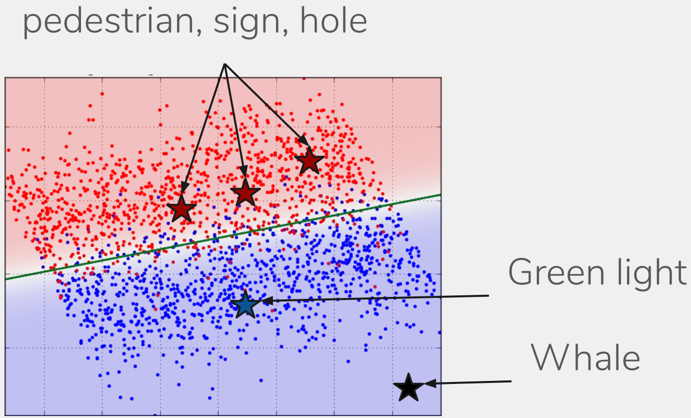




Exploring beyond the known frontiers

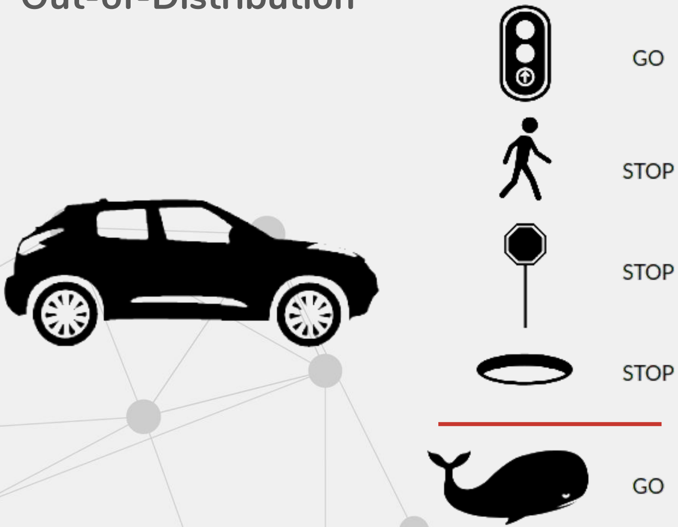
- **Uncertainty Estimation**
- 3 types of uncertainty:
    - Aleatoric
    - Epistemic
    - **Out-of-Distribution**



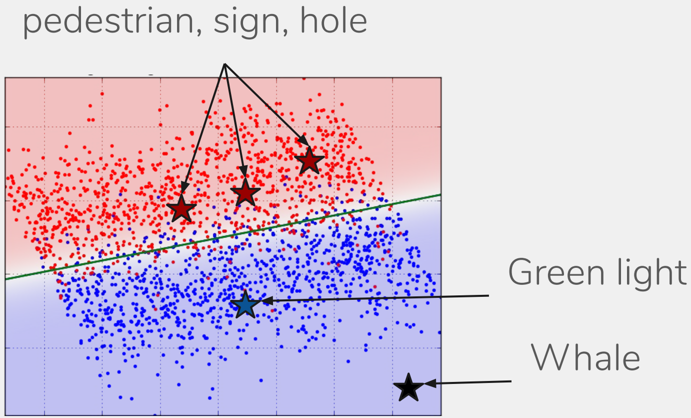


Exploring beyond the known frontiers

- **Uncertainty Estimation**
- 3 types of uncertainty:
    - Aleatoric
    - Epistemic
    - **Out-of-Distribution**



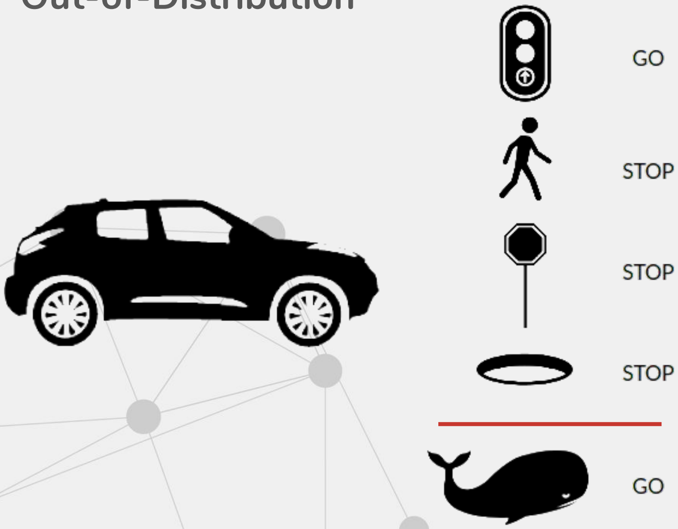
Rule #1 in ML: Training and Test come from the same distribution

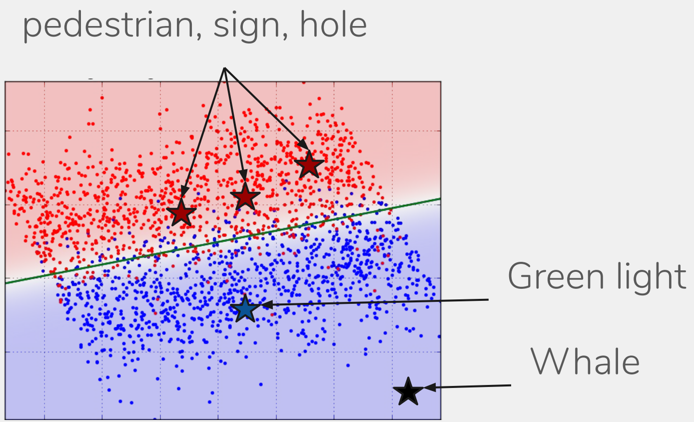


Exploring beyond the known frontiers

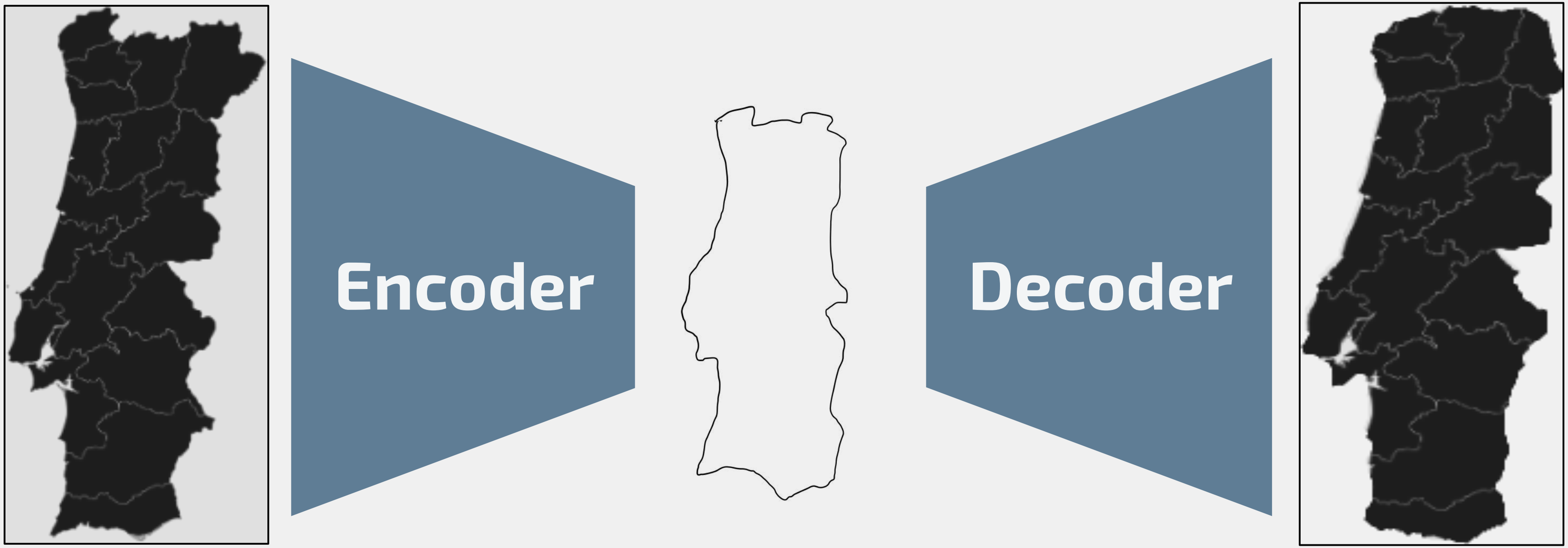
- **Uncertainty Estimation**
- 3 types of uncertainty:
  - Aleatoric
  - Epistemic
  - **Out-of-Distribution**

~~**Rule #1 in ML: Training and Test come from the same distribution**~~

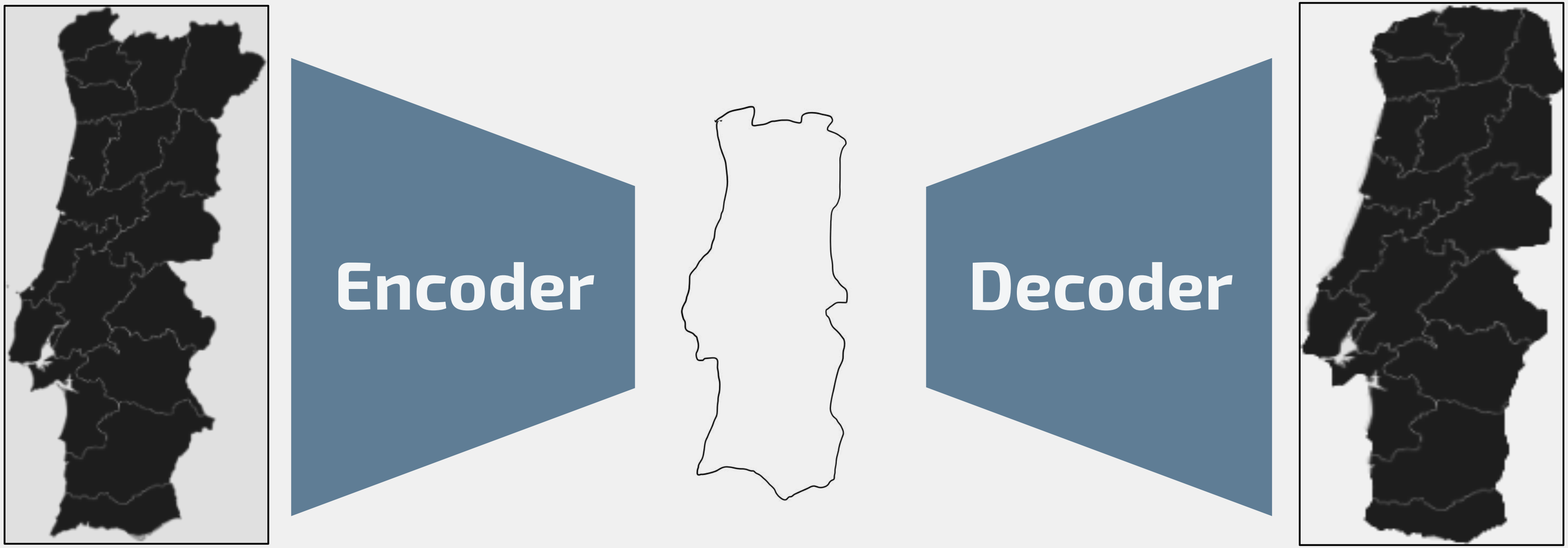




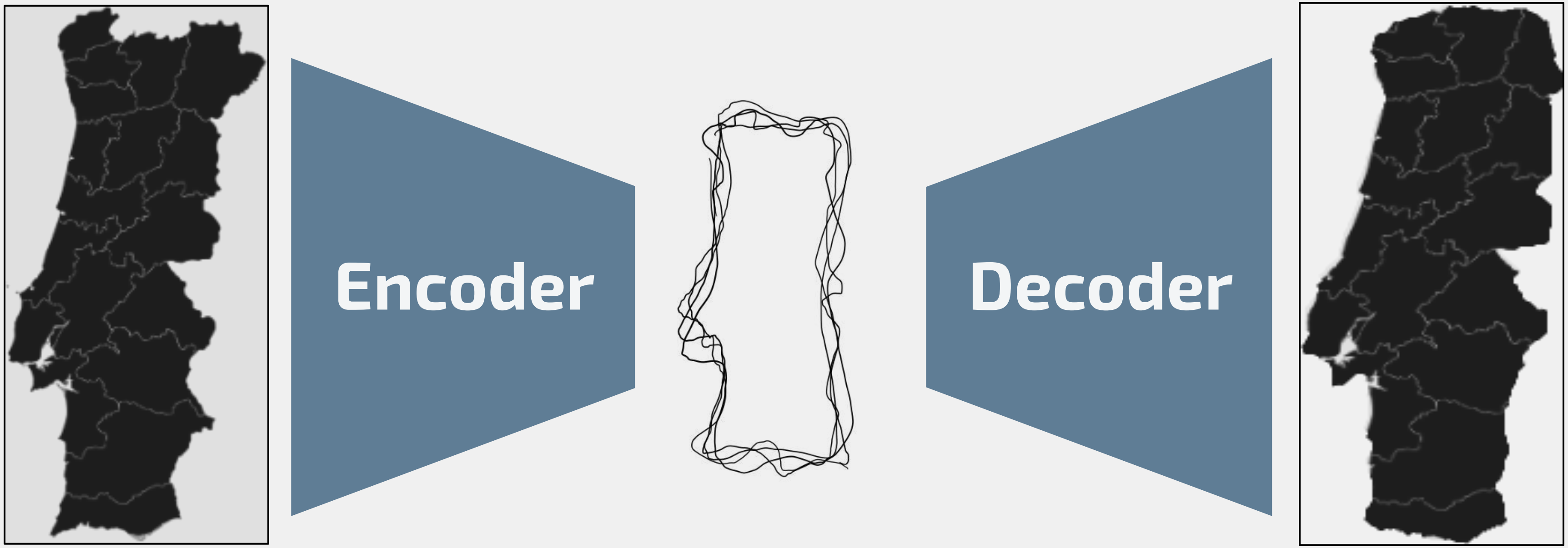
Variational Auto-Encoders



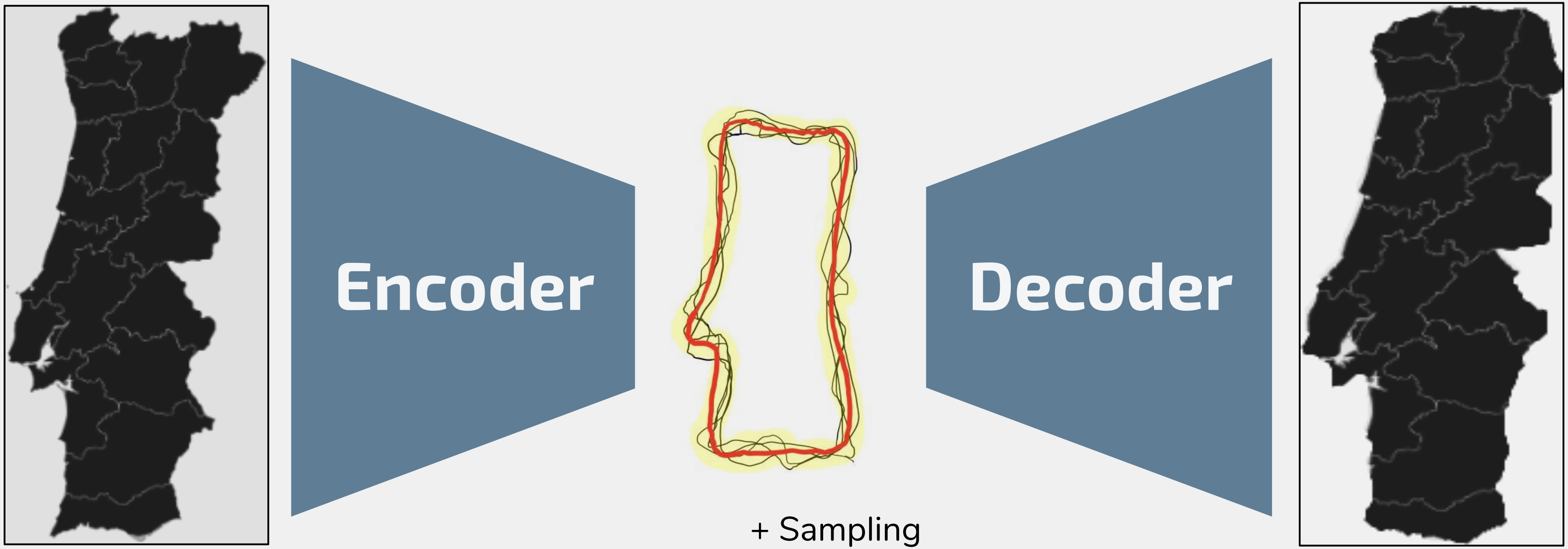
Variational Auto-Encoders



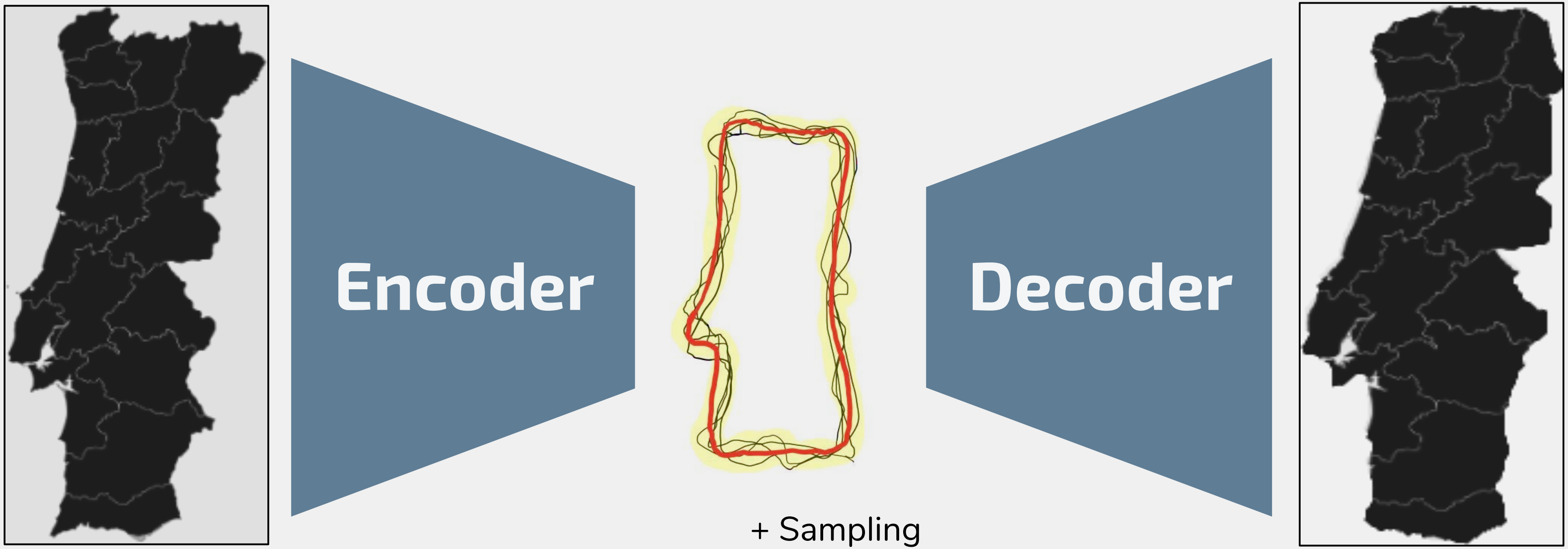
Variational Auto-Encoders



Variational Auto-Encoders



Variational Auto-Encoders



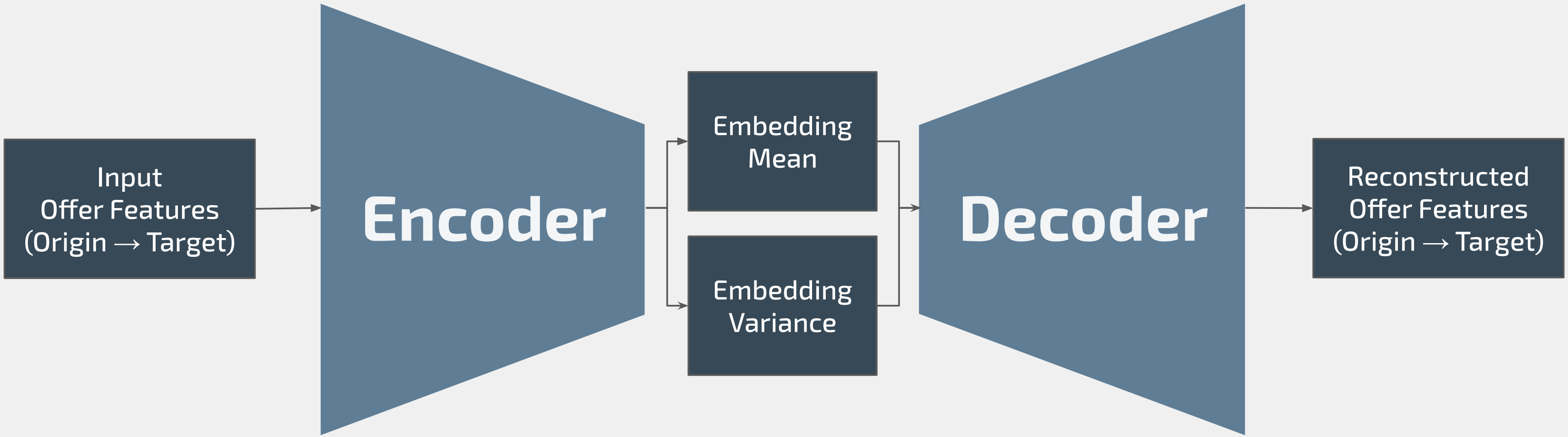
- **Relevant regions are well-defined:**
  - Very well compressed with all the relevant information...
  - We know how to reconstruct with low margin of error

Variational Auto-Encoders

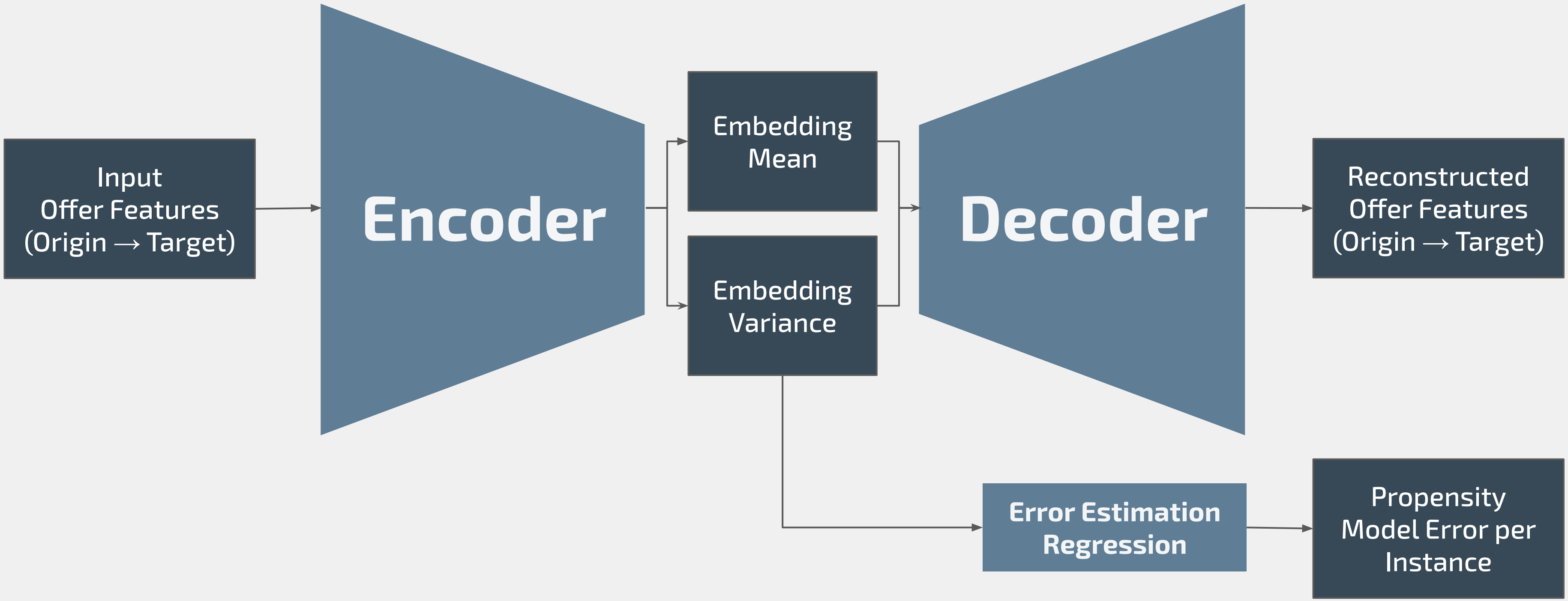
+ Sampling

- **Unexplored regions are loosely-defined:**
  - Kind of know how they work but we lost some information…
  - we don't know how to reconstruct

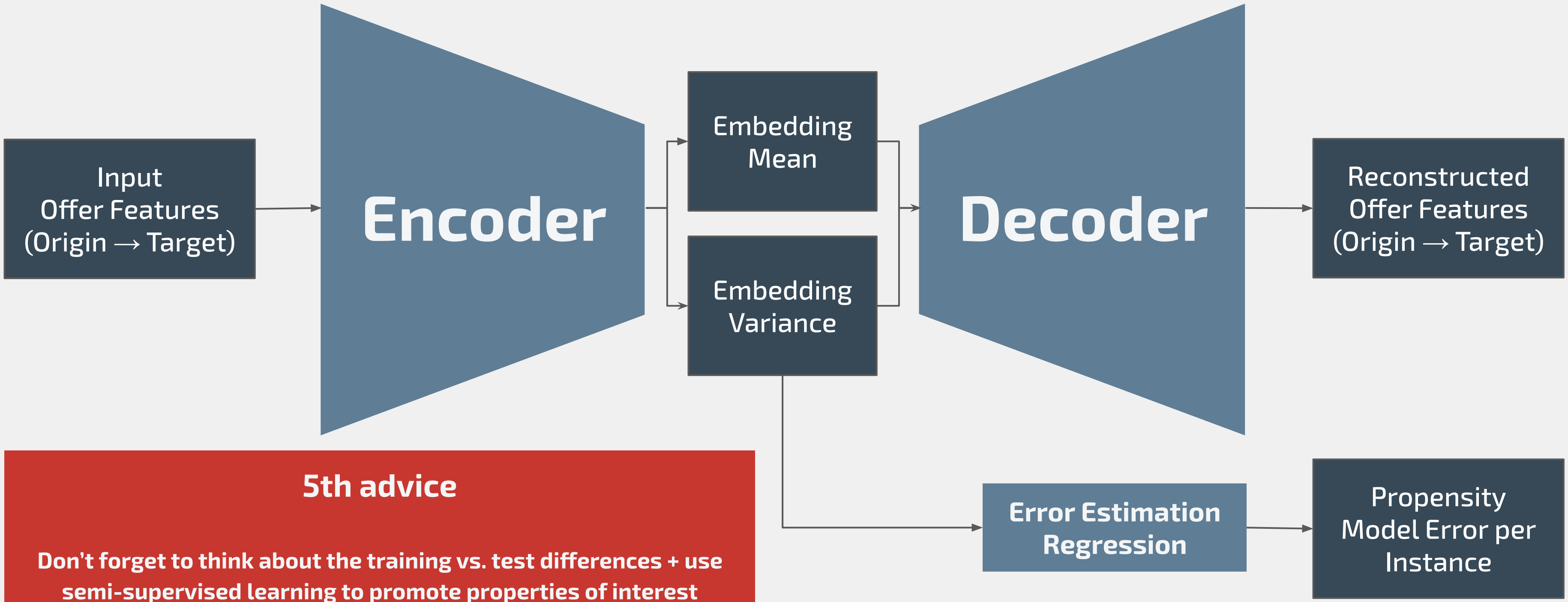
Variational Auto-Encoders



Variational Auto-Encoders + Multitask



Variational Auto-Encoders + Multitask

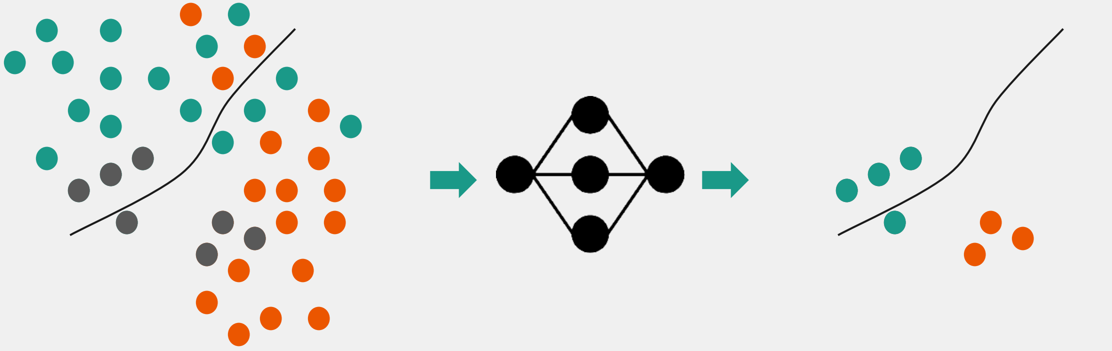


5th advice

Don't forget to think about the training vs. test differences + use semi-supervised learning to promote properties of interest

in partnership with

Some Results



Hide data

Train both models

Errors vs. uncertainty

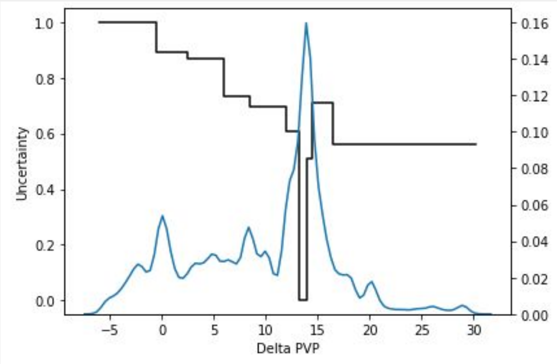
Spearman's Rank:

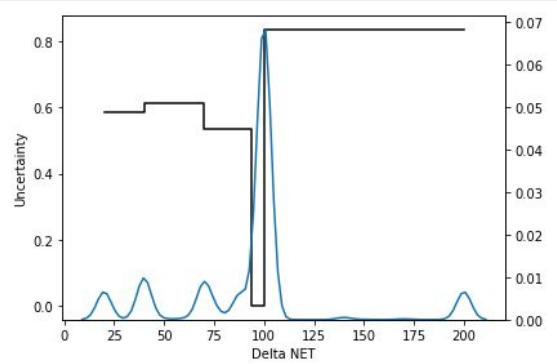
Hard tests:

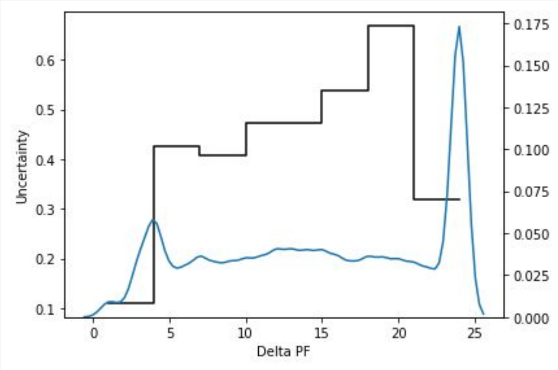
0.20

Soft tests:

0.45







Current Work

- **Goal 1:** How to ensure by construction that **monotonic features** have **monotonic impact** on propensity?
  - e.g., as price increases, probability decreases
  - e.g., as the data plafond increases, the probability increases

- **Goal 2:** can we learn more than one task simultaneously?
  - Multi-task Learning
  - Joint NBO, Churn, CLV, Channel propensity
  - Customer Journey problem

Main Learning Points

1st advice

Cold Start + Domain Knowledge ⇒ Content-based RecSys

2nd advice

Sometimes dividing doesn't mean conquering

3rd advice

Merge alternative data sources related to your target task

(transfer learning isn't just weight transfer + fine-tuning)

4th advice

Calibrate your probabilities in any decision-making process, especially if you manipulated the distributions!

5th advice

Don't forget to think about the training vs. test differences + use semi-supervised learning to promote properties of interest

Main Learning Points

nilg.ai in partnership with NOS

1st advice

Cold Start + Domain Knowledge ⇒ Content-based RecSys

2nd advice

Sometimes dividing doesn't mean conquering

3rd advice

Merge alternative data sources related to your target task

(transfer learning isn't just weight transfer + fine-tuning)

4th advice

Calibrate your probabilities in any decision-making process, especially if you manipulated the distributions!

5th advice

Don't forget to think about the training vs. test differences + use semi-supervised learning to promote properties of interest

Bonus advice:

Go beyond classification and regression. Learn more paradigms so you can choose a better hammer!

Providing relevant recommendations beyond the explored frontiers

Kelwin Fernandes, Ph.D.
CEO @ NILG.AI
20191203

Acknowledgments and Resources

- NOS Data Science Team
- Data Science Portugal
- **Images from:**
  - Eirik Solheim @ Flickr
  - Pexels
- **Icons:** flaticon