© Raquel H. Ribeiro

Exposing fraud with machine learning

DS Portugal Meetup @ UPtec
Porto, 24th May 2017

Roadmap

1. A payments world overview
2. Where there is value, there is fraud
3. Using machine learning to uncover and prevent fraud

About Feedzai

- Young start-up with a Science & Engineering DNA
- Expertise in forecasting and fraud detection
- Processes transactions worth more than Portugal’s GDP

About DS team

- Currently over 25 data scientists from various backgrounds in Lisbon, Coimbra, Porto, New York and Atlanta - yet still growing!
- Delivery team works in PoCs and established projects, from banking to e-commerce
- Research team leads innovation
- Code review and reading groups are standard practice

The payments world

the customer
the merchant
the acquirer
the issuer

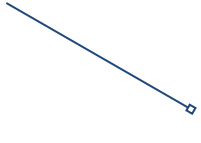
How to buy services and goods?

card present

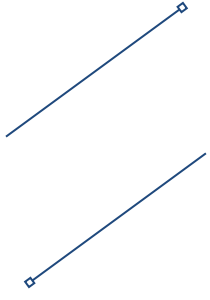
card not present

Customer

In a point of sale near you



terminal



Merchant

issuer

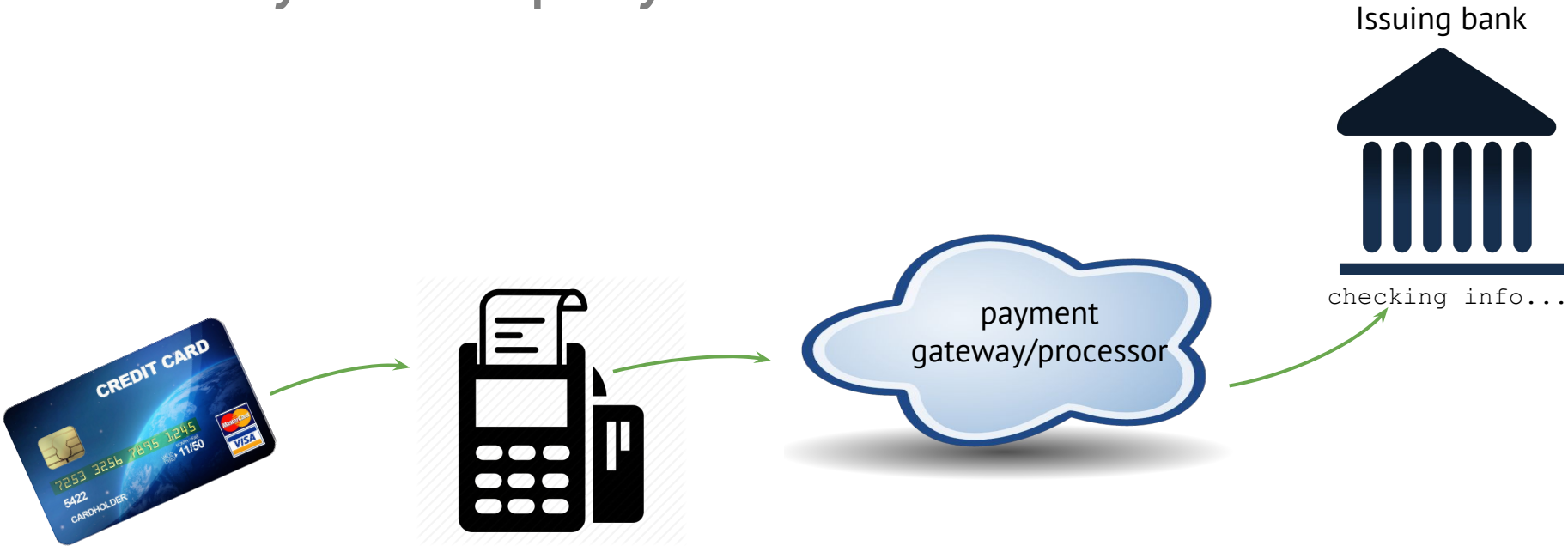
acquirer

When you swipe your card

When you swipe your card

When you swipe your card

When you swipe your card



When you swipe your card

Issuing bank

When you swipe your card

Issuing bank

When you swipe your card

Issuing bank

Fraudster enters the play

Fraudster enters the play

...	...	...
7253 3256 7845 1245	12/17	055
...	...	...

Fraudster enters the play

...	...	...
7253 3256 7845 1245	12/17	055
...	...	...

Fraudster enters the play

...	...	...
7253 3256 7845 1245	12/17	055
...	...	...

Fraudster enters the play

Issuing bank

...	...	...
7253 3256 7845 1245	12/17	055
...	...	...

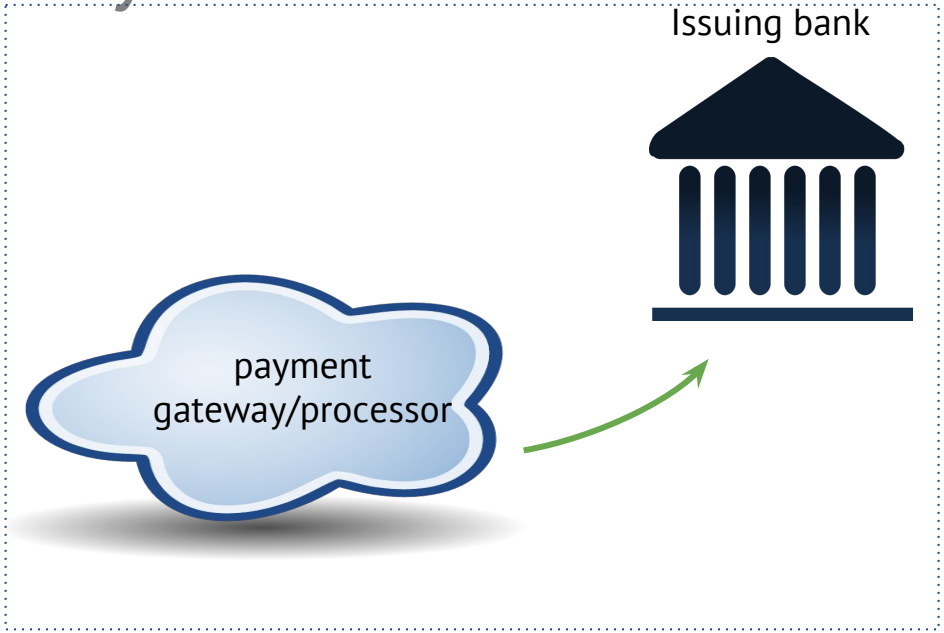
Fraudster enters the play

...	...	...
7253 3256 7845 1245	12/17	055
...	...	...

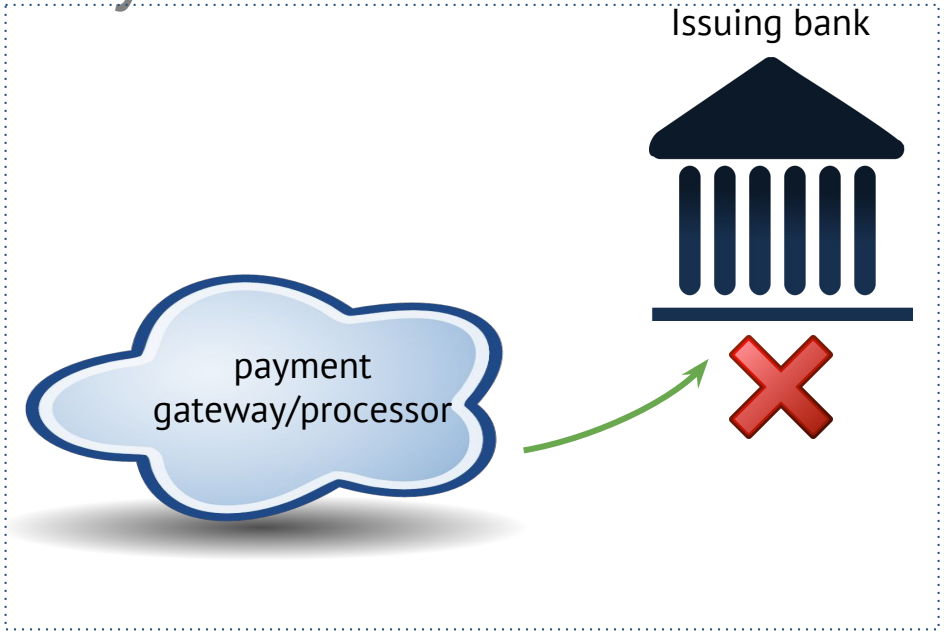
Issuing bank

Fraudster enters the play

...	...	...
7253 3256 7845 1245	12/17	055
...	...	...



Fraudster enters the play



...	...	...
7253 3256 7845 1245	12/17	055
...	...	...

But I protect my card!

“Where there is value, there is fraud”

data breaches

the fraud landscape

ATM machines

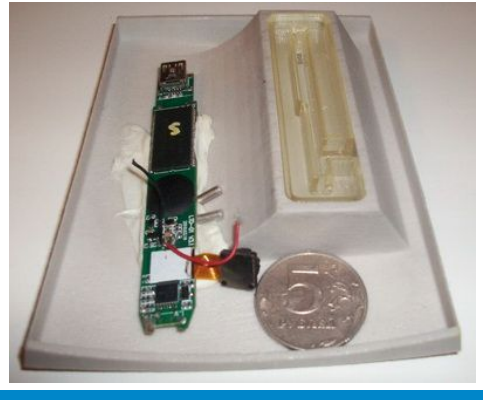
Well, at least one is an ATM machine…!

This is a card skimming device!

before

after

What happened? Mini card scanner…

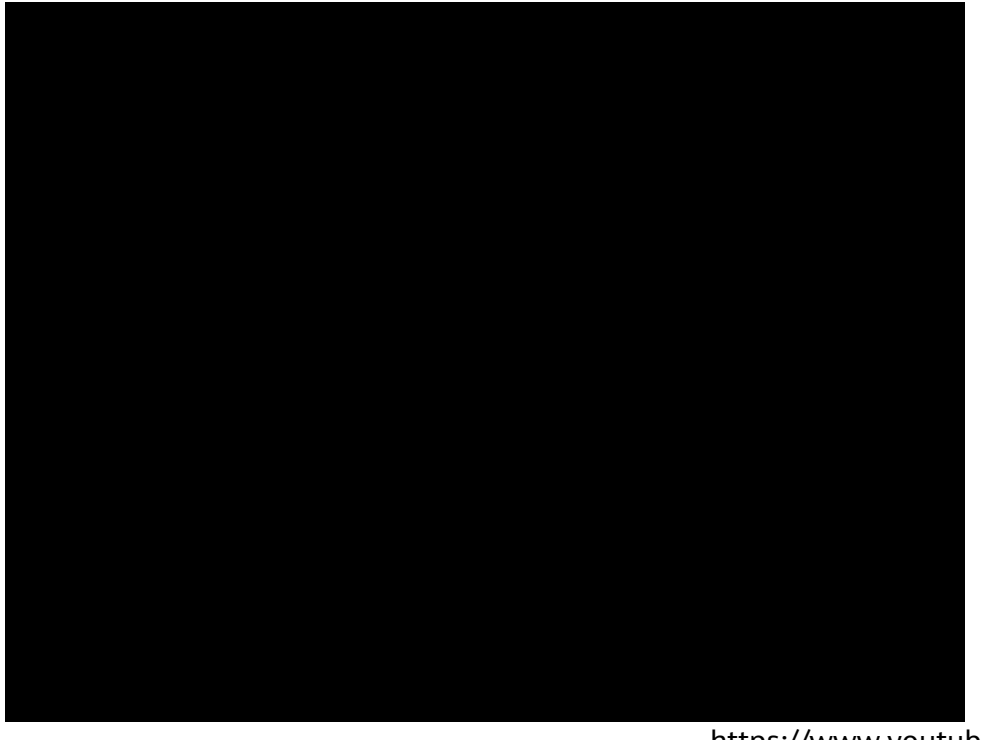


..complete with camera and fake keyboard

If you avoid ATMs…

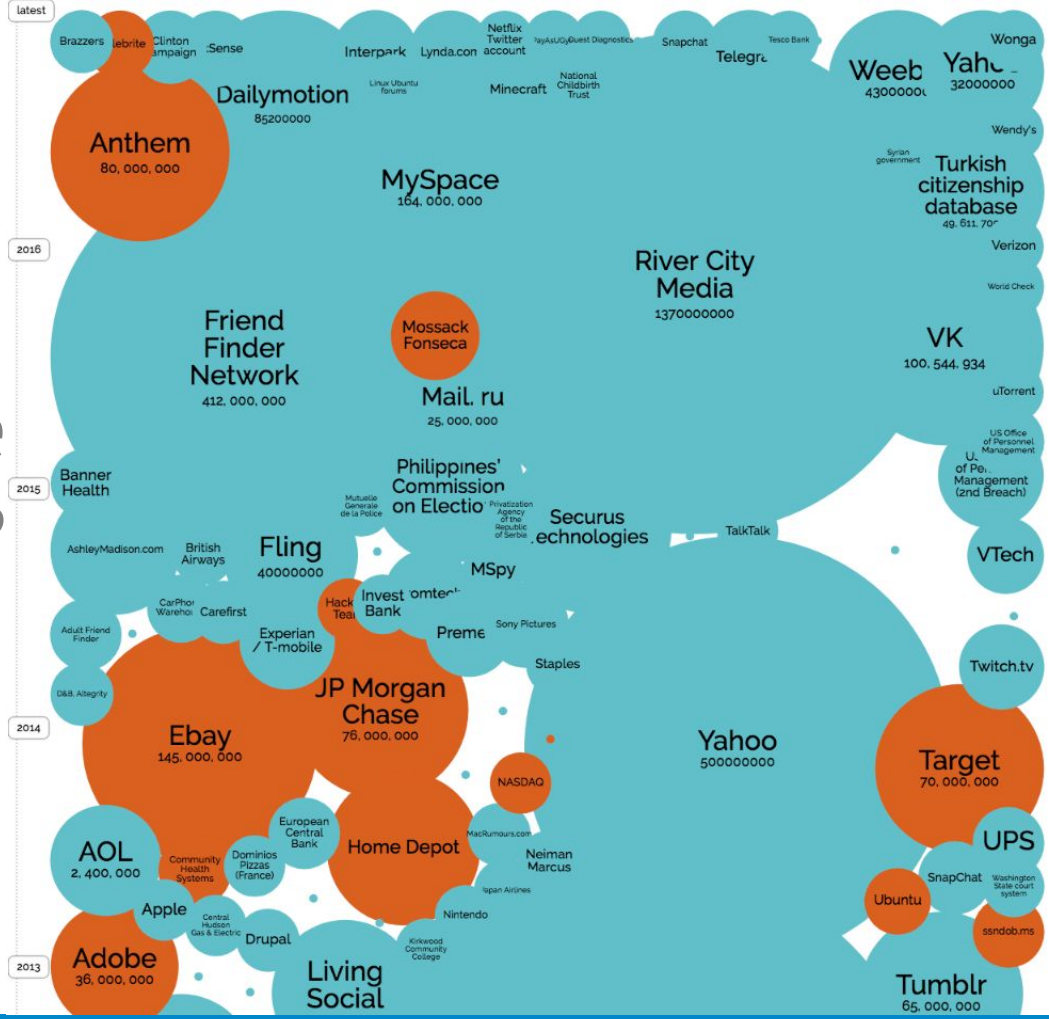
…you can still be targeted

…at an astonishing speed



https://www.youtube.com/watch?v=y83ZgzuFBSE

How serious is the data breach landscape?



From *Information is Beautiful*
https://goo.gl/wL3hEv

If my card is targeted, who is liable?

Issuer

Merchant

card present

card not present

Customer

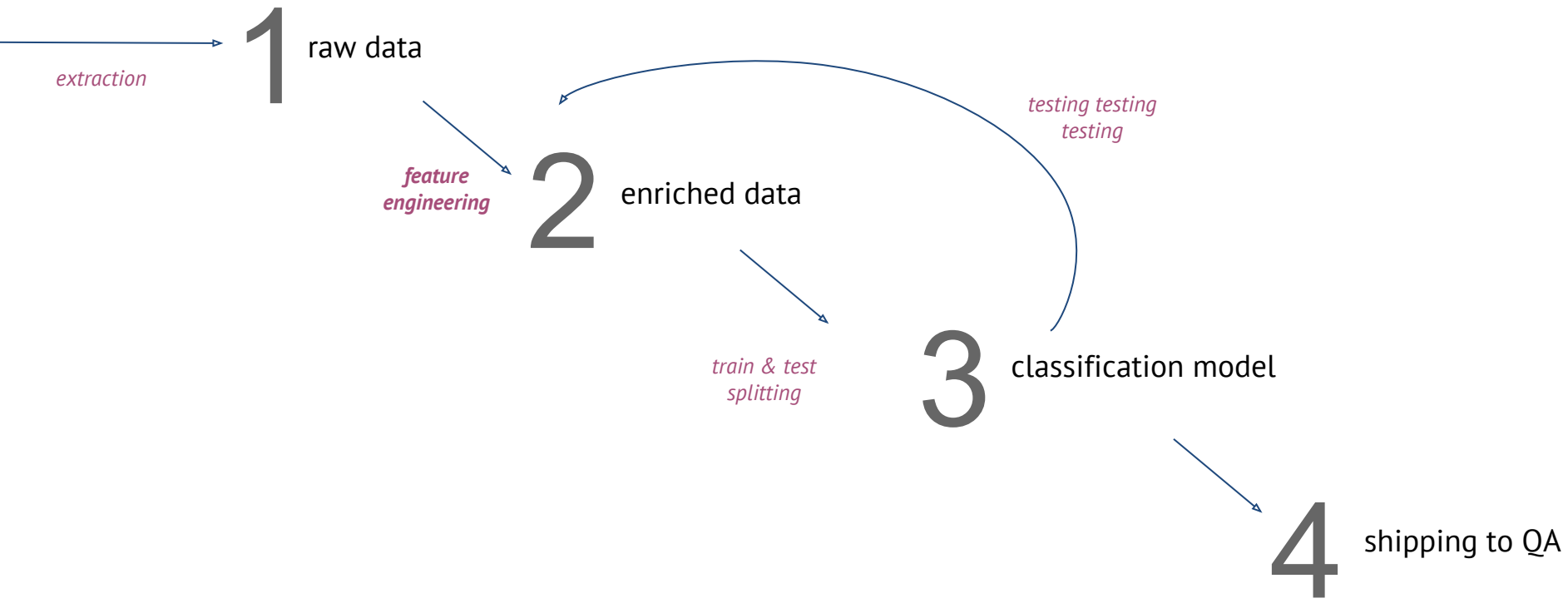
Machine learning insights

A - do you trust your labels?

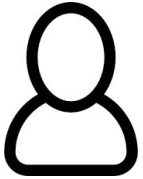
B - how do you evaluate your model?

problems you won’t usually encounter in kaggle competitions

A Data Science pipeline



A - When will you know it was fraud?



- You call your bank and request a chargeback
- You are reimbursed - happy ending!

Merchant is notified to pay the bill

30 days after transaction happened

FACT: the classification model failed to identify a fraudulent transaction

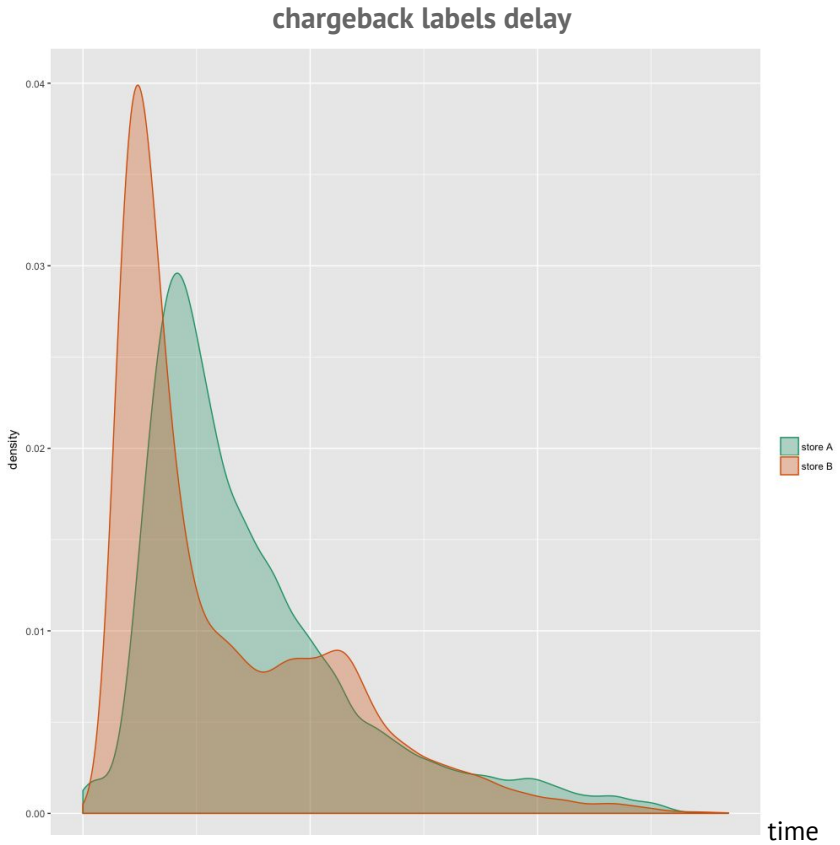
PROBLEM: has the wrong label been fed into the model already?

A - Do you trust your labels?

How long are you willing to wait to get the purest labels?

The answer is a compromise between

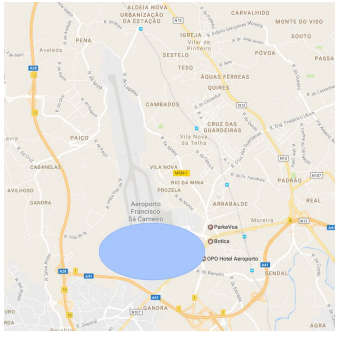
more recent data and the correct labels.



B - Tell me your metric

We deal with unbalanced datasets that range from

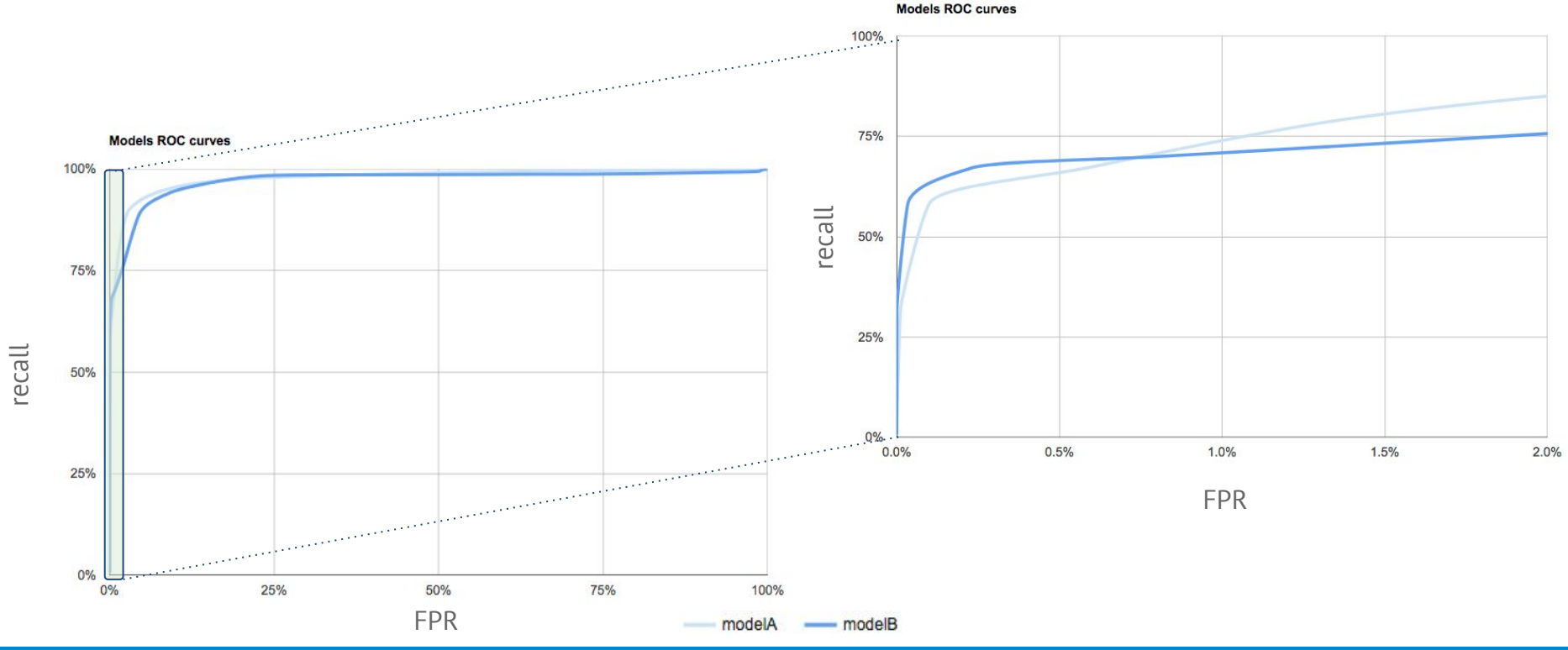
1:10 to 1:100000



A dummy classifier can reach over 90% in accuracy and miss all the fraud!

B - Decoding ROC curves

synthetic data



B - Scoring transactions in e-commerce

Though this is a classification problem, our models produce a score: from 0 to 1000.

Score 930 probably fraud send to review

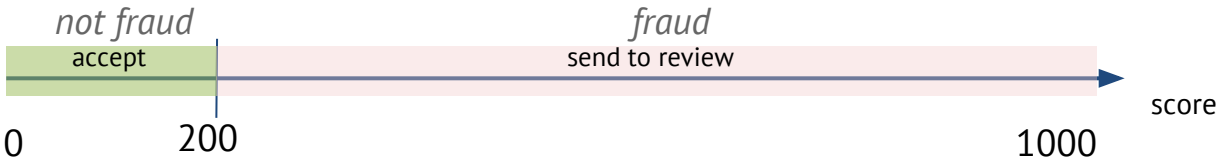
Score 20 probably legit approve

How do you define the threshold for fraud and not fraud?

B - How would you select a score threshold?

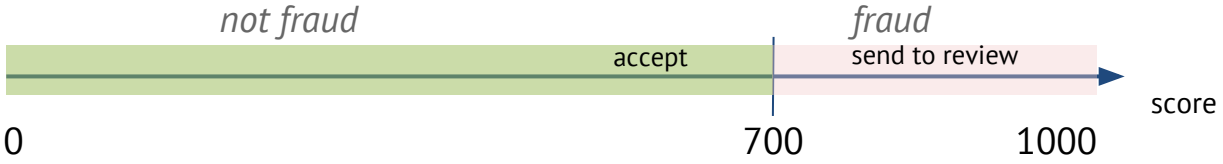
Though this is a classification problem, our models produce a score:

A



Catches most fraud but also many legitimate cases

B

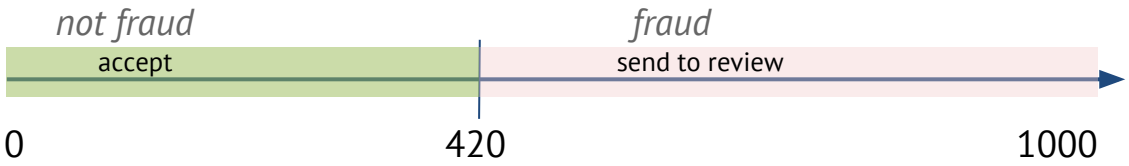


Fails to catch some fraud but should trigger less FPs

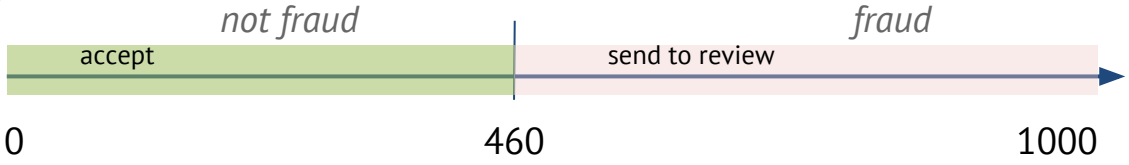
B - How would you select a score threshold?

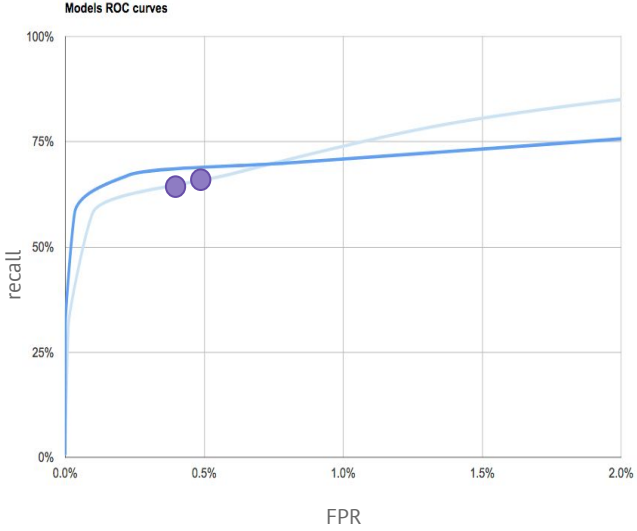
But how do you decide between very close thresholds?

A



B





B - How do you evaluate your model?

Different thresholds will work for different SLAs.

Do you know what is the cost of a **FP** compared to a **TP**?

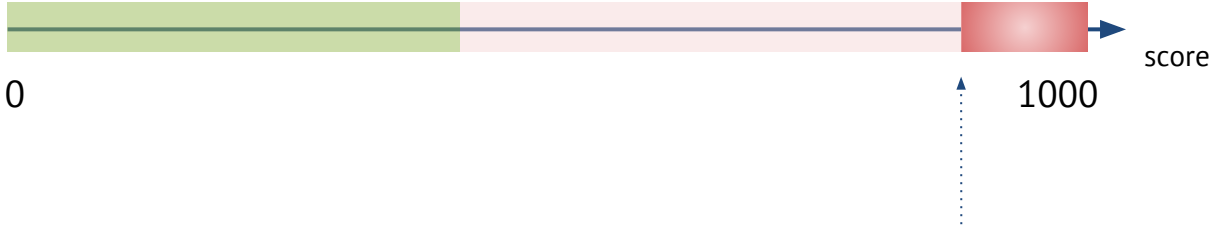
We can use **recall** to determine how much fraud we aim at catching and **FPR** to adjust how many FPs we are willing to allow for, cf.:

$$\text{recall} = \frac{TP}{TP + FN} \quad \text{and} \quad \text{FPR} = \frac{FP}{FP + TN}$$

Balance between catching fraud and keeping good customers satisfied with the service

B - How do you evaluate your model?

For those transactions which you are pretty certain are fraud, can you help the review process?



Do you still want to use **recall** to determine this threshold?

Are other metrics better suited for this? $\text{precision} = \frac{TP}{TP + FP}$

Machine Learning wrap-up

Summary

- Machine learning models face a spectacular adversarial problem in fighting fraud. Feature engineering and rapid deployments are key for success.

- In this talk we discussed two real-world problems:
  - in the business world, the labels arrive late and your models need to be deployed fast
  - different metrics for the algorithm performance are best suited for different objectives

- **Fighting fraud is very challenging but intellectually very stimulating!**

Thank-you!

Raquel H Ribeiro
Raquel.Ribeiro@feedzai.com

feedzai