NYU FRE 7773 - Week 5

Machine Learning in Financial Engineering

Ethan Rosenthal

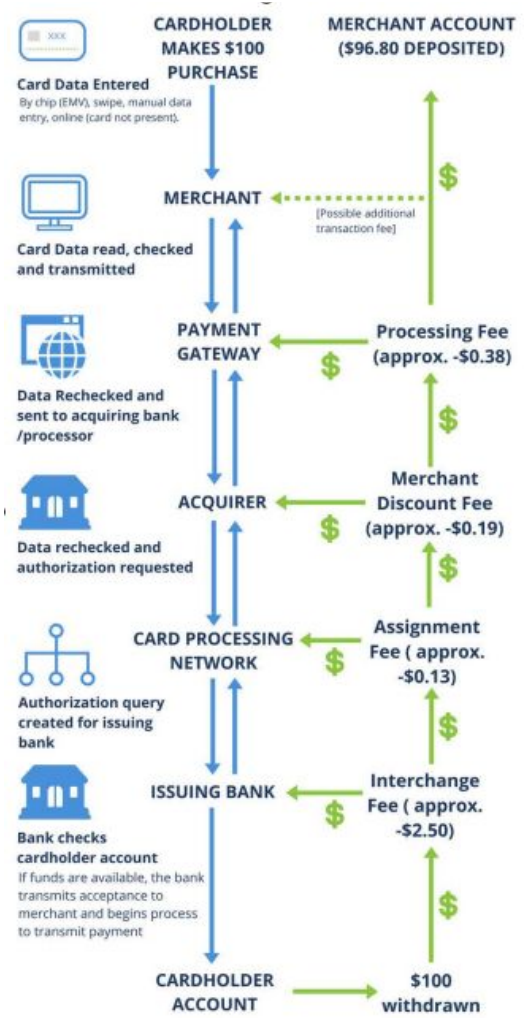
Case Study: Fraud Detection

Machine Learning in Financial Engineering

Ethan Rosenthal

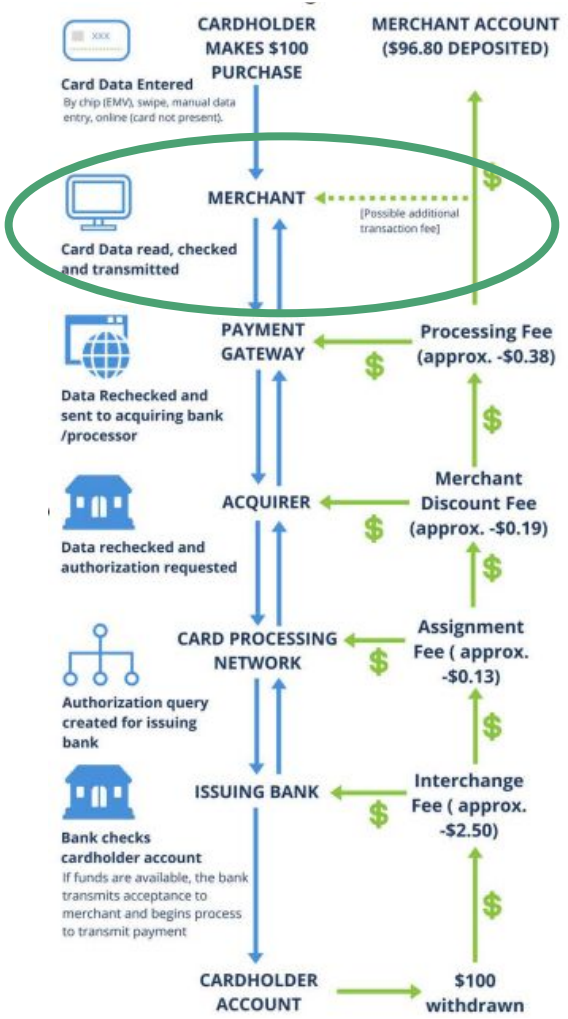
Risk

Square



https://go.wepay.com/collateral/payment-processing-basics/

Square

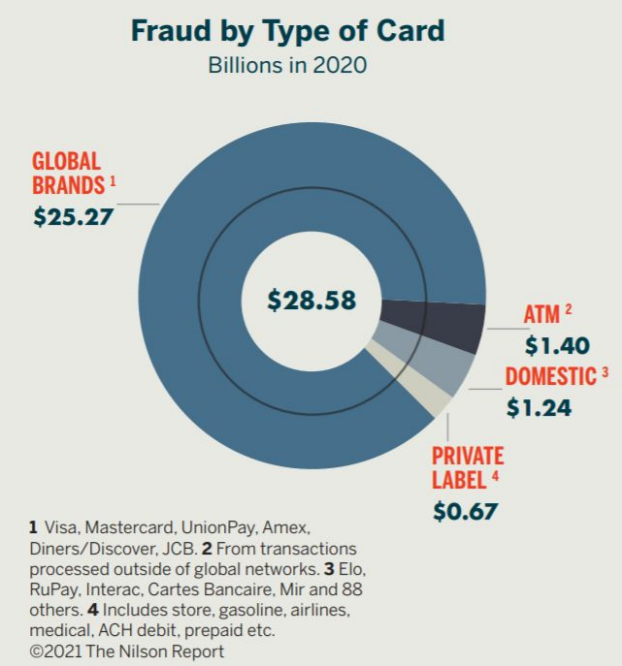


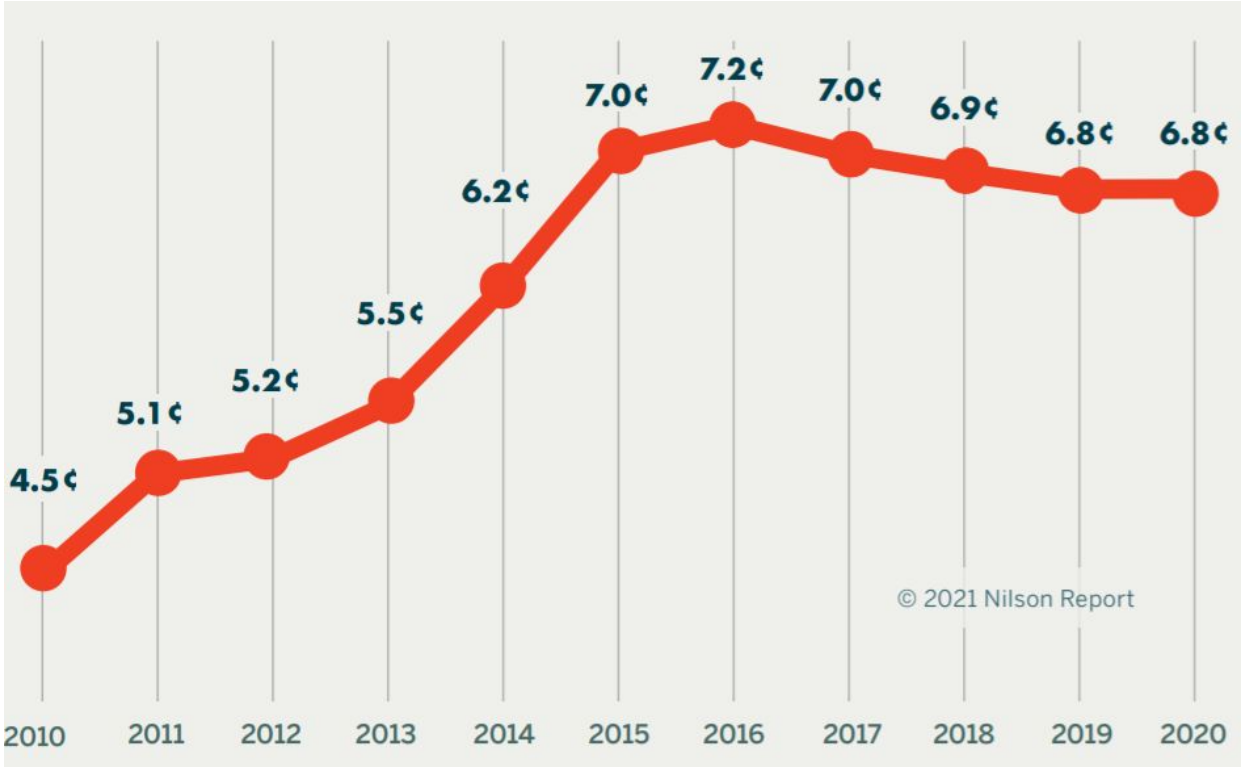
https://go.wepay.com/collateral/payment-processing-basics/

Square Launches On-Demand and Instant Payments for Square Payroll Customers

Square introduced two new features this week to help its customers using Square Payroll – as well as their employees – manage their cash flow.

Fraud





Nilson Report Issue 1209 https://nilsonreport.com/upload/content_promo/NilsonReport_Issue1209.pdf

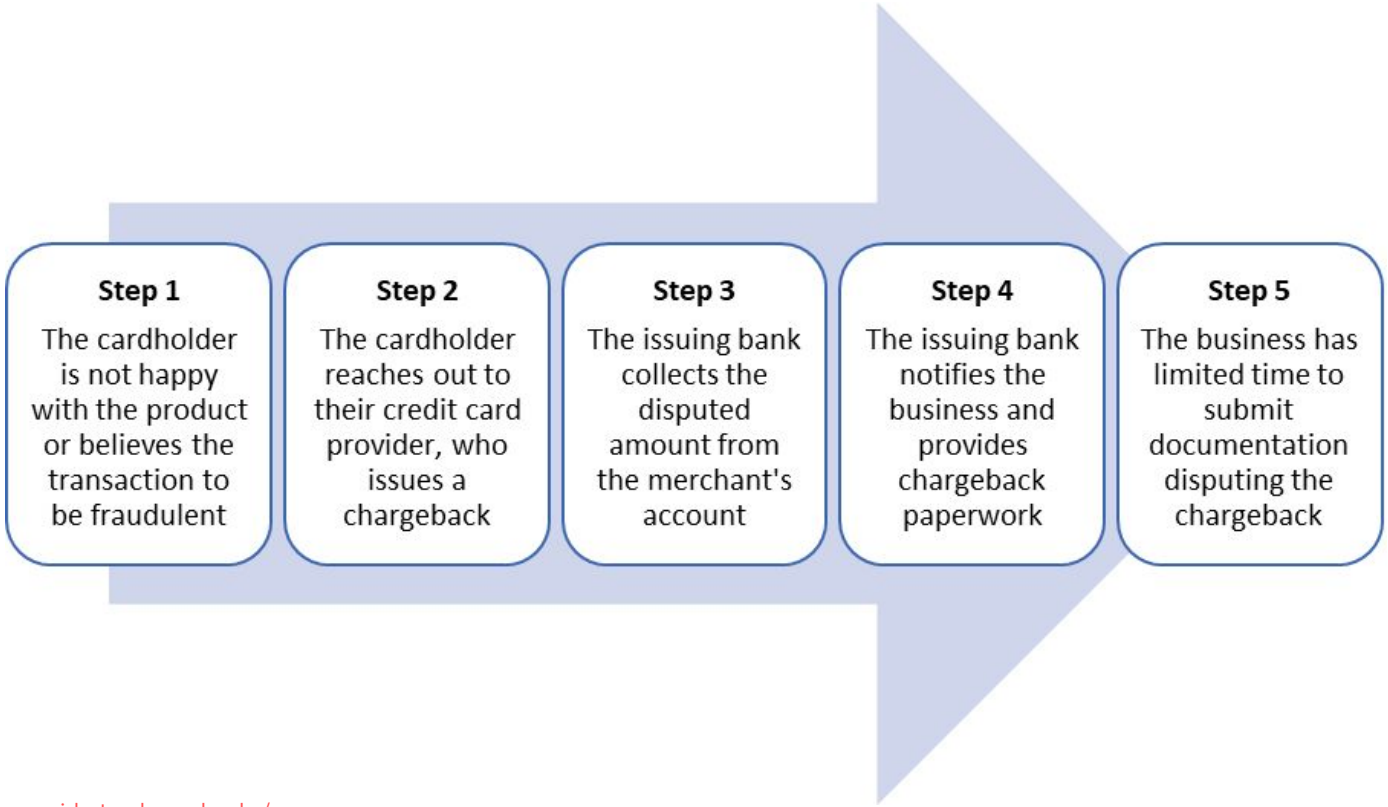
Card Fraud Projected through 2030

YEAR	Total Volume (TRIL.)	Fraud (BIL.)	Cents per $100 VOLUME
2020	$41.962	$28.58	6.81
2021	$47.229	$32.20	6.82
2022	$50.868	$34.36	6.75
2023	$54.061	$36.13	6.68
2024	$57.323	$38.07	6.64
2025	$60.583	$39.89	6.58
2026	$64.038	$41.73	6.52
2027	$67.570	$43.76	6.48
2028	$71.221	$45.54	6.39
2029	$75.111	$47.50	6.32
2030	$79.140	$49.32	6.23

© 2021 Nilson Report

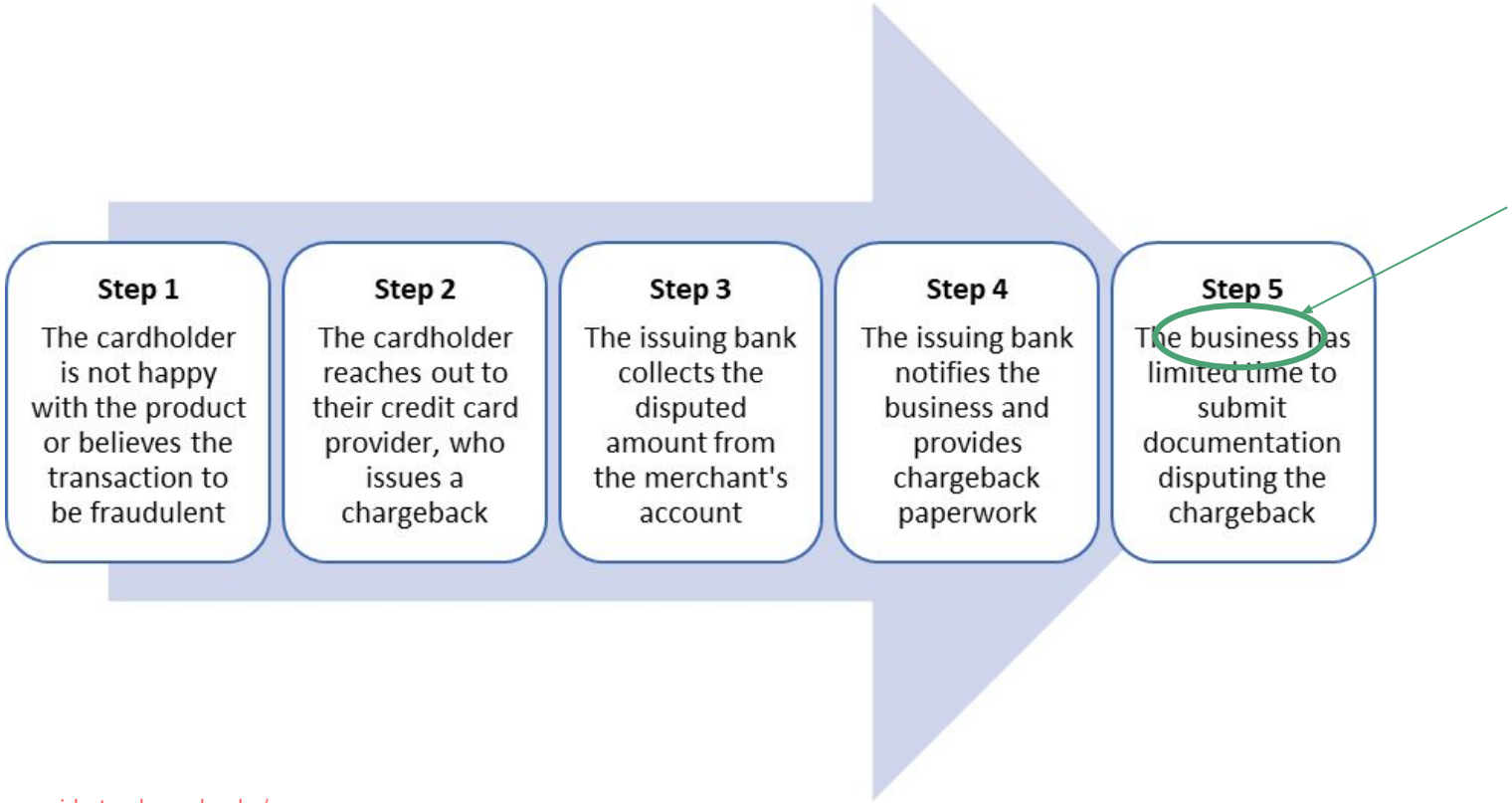
Nilson Report Issue 1209 https://nilsonreport.com/upload/content_promo/NilsonReport_Issue1209.pdf

Chargebacks



https://basyspro.com/your-guide-to-chargebacks/

Chargebacks

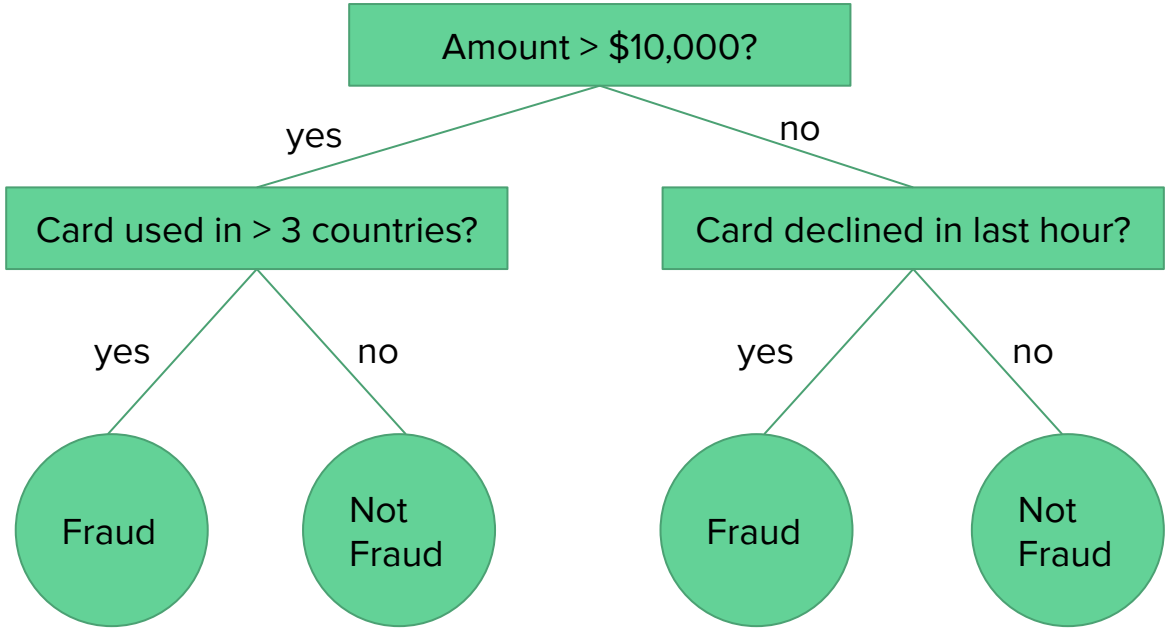


https://basyspro.com/your-guide-to-chargebacks/

Fraud Detection

Rules

- A manual decision tree
- Ideally, high precision, low recall
- Quick and dirty.
- Lots of manual work and tuning.

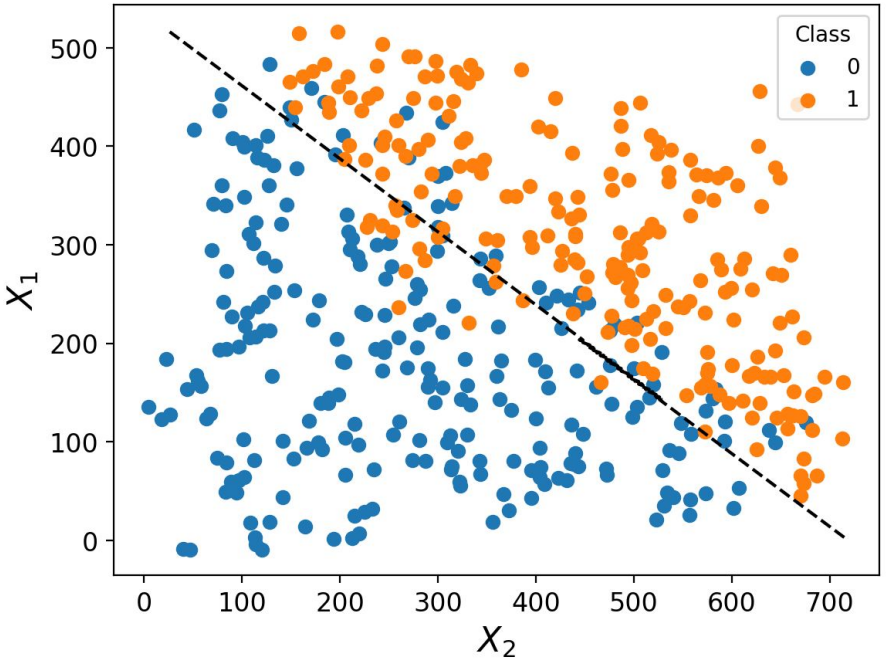


Machine Learning Models

- Binary classification: fraud vs. not fraud
- Sample -> often an **event** (e.g. payment transaction, bank withdrawal).
- **X** -> event features + auxiliary features
- **y** -> was event “associated” with fraud or not
- Train model on historical events, predict on real time events.

Machine Learning Models

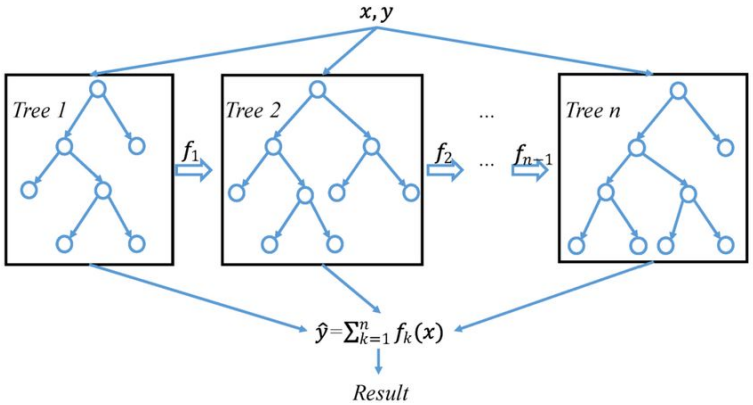
- Logistic regression is simple and can be extremely fast for low latency applications.



Machine Learning Models

- Tree-based models work particularly well.
- Domain naturally lends itself to if/else statements.
- Easy to tune.
- Relatively fast to train and predict.
- Somewhat interpretable.

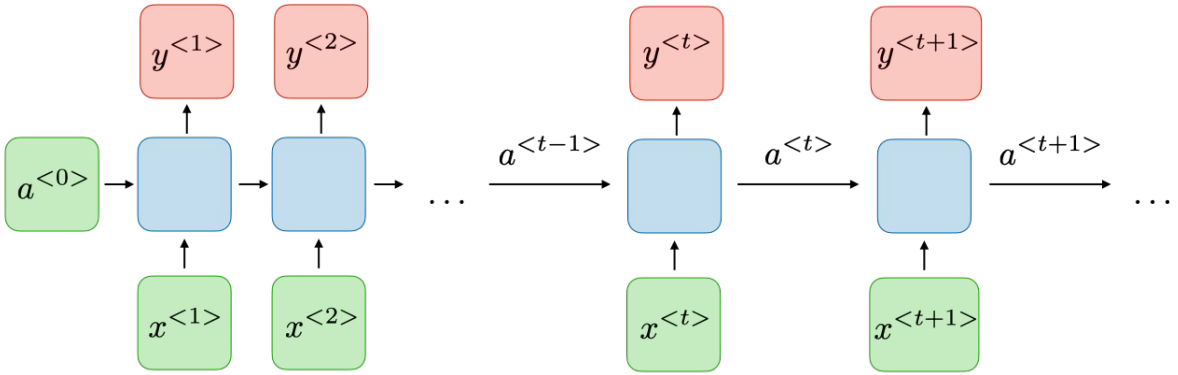
dmlc XGBoost



https://link.springer.com/article/10.1007/s10509-019-3602-4

Machine Learning Models

- Deep learning models can handle sequences of events.
- Much trickier to model and tune.



TensorFlow

PyTorch

https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks

Fraud Detection Difficulties


KEANU REEVES DENNIS HOPPER SANDRA BULLOCK
GET READY FOR RUSH HOUR.
SPEED
JUNE 10

Speed – Realtime Features

- Payment Amount
- Transaction Country
- Credit Card Brand
- IP Address
- Tip Percentage

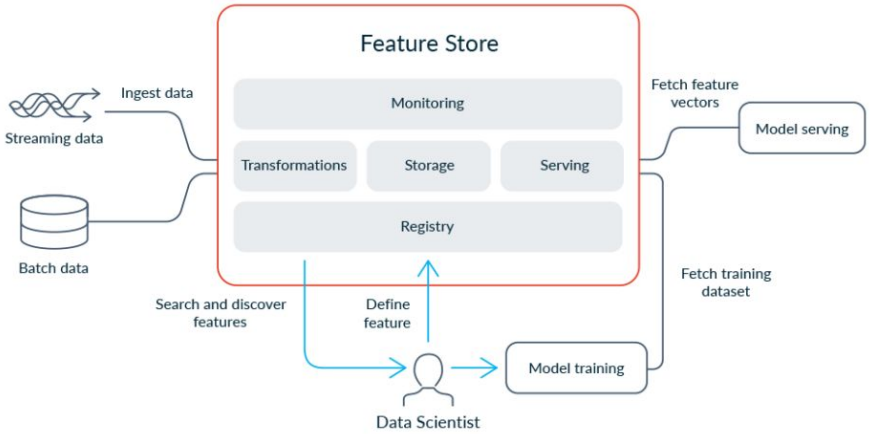
Speed – Realtime Features

- Standard deviation of milliseconds between transaction attempts over the last 30 seconds.

Speed – Realtime Features

Feature Stores are systems for ingesting streams of events and converting them into features, *quickly*.

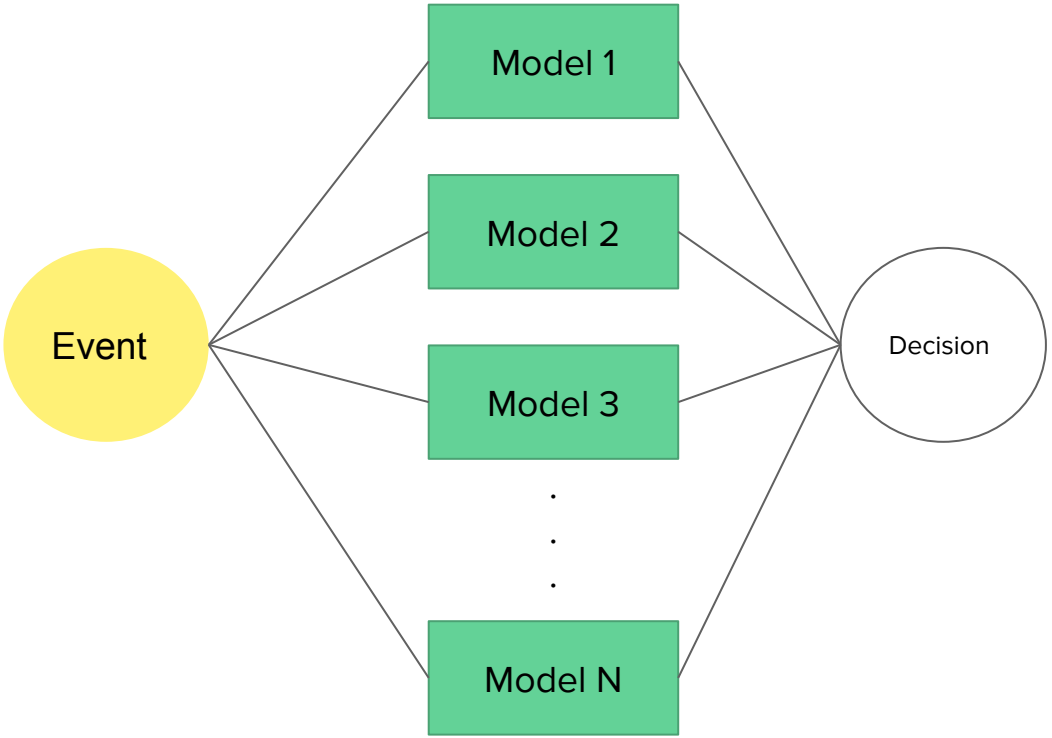
Tradeoff between feature *accuracy* and *latency*.



tecton

Speed – Realtime Detection

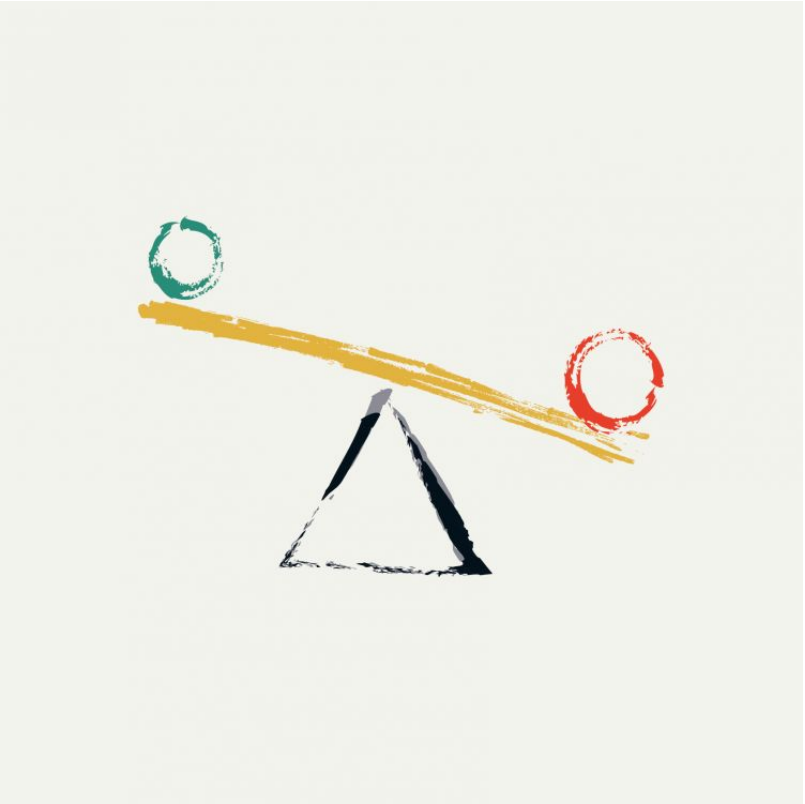
- Thousands or rules and models for a given event.
- Limited time that each can take.
- Sync vs. async workflows.



Fraud Decisioning

Tradeoffs

- Automation vs. Confidence
  - Block high fraud likelihood transactions.
  - Manually review when not sure.
- Friction vs. Risk
  - Verify via SMS vs. providing bank documents.
- The goal is to “cost more” than the alternative.
- **False Positives vs. False Negatives**



lil du bois
@_lildubois

well, well, well, if it isn't the consequences of my own actions

1:25 AM · Sep 29, 2018 · Twitter for Android

133.7K Retweets **5,164** Quote Tweets **297K** Likes

Training Models

Machine Learning Models

- Binary classification: fraud vs. not fraud
- Sample -> often an **event** (e.g. payment transaction, bank withdrawal).
- **X** -> event features + auxiliary features
- **y** -> was event “associated” with fraud or not
- Train model on historical events, predict on real time events.

Machine Learning Models

- Binary classification: fraud vs. not fraud
- Sample -> often an **event** (e.g. payment transaction, bank withdrawal).
- **X** -> event features + auxiliary features
- **y** -> was event “associated” with fraud or not
- Train model on historical events, predict on real time events.

Speed – ~~Realtime~~ Historical Features

- Standard deviation of milliseconds between transaction attempts over the last 30 seconds at the time the Fraud model would make its prediction.

Speed – ~~Realtime~~ Historical Features

- Standard deviation of milliseconds between transaction attempts over the last 30 seconds at the time the Fraud model would make its prediction.
- Don’t leak the future into the past!

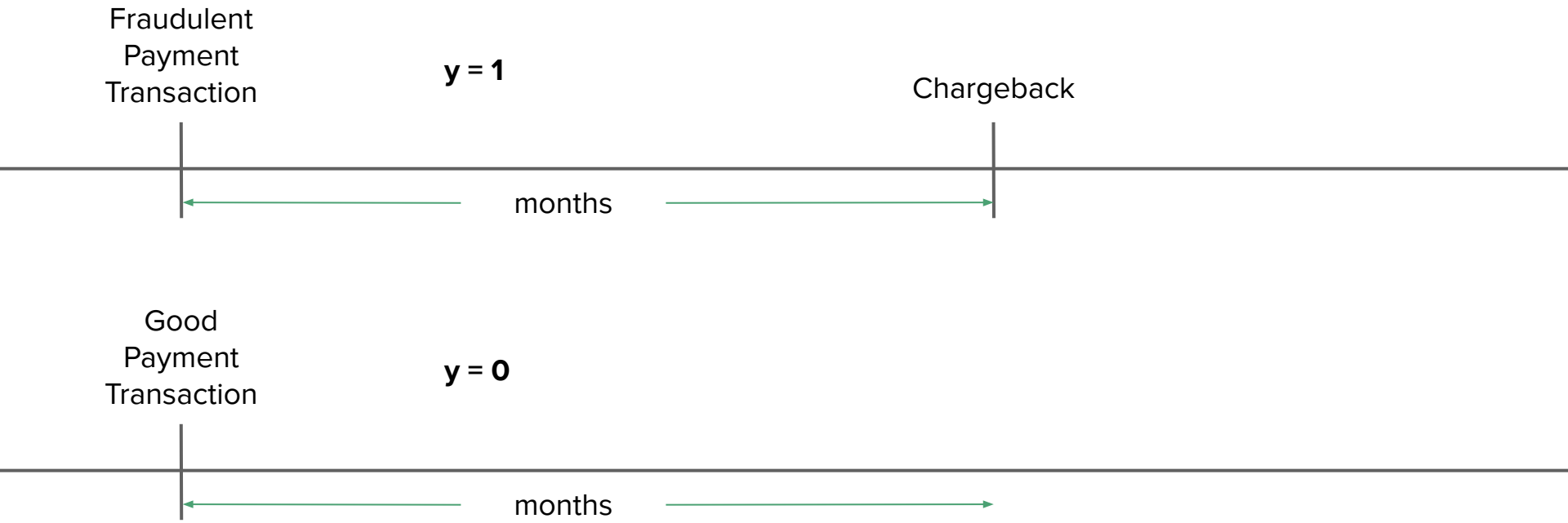
Speed – ~~Realtime~~ Historical Features

- Standard deviation of milliseconds between transaction attempts over the last 30 seconds at the time the Fraud model would make its prediction.
- Don’t leak the future into the past!
- Historical features must match realtime features (“train/test skew”)

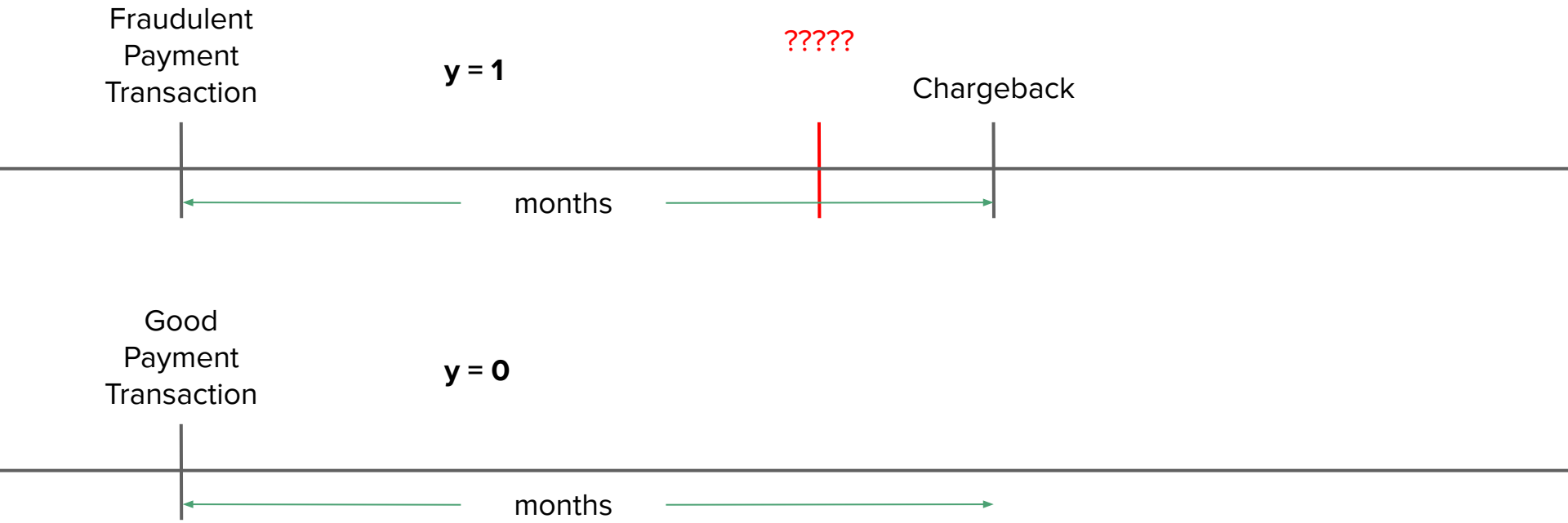
Machine Learning Models

- Binary classification: fraud vs. not fraud
- Sample -> often an **event** (e.g. payment transaction, bank withdrawal).
- **X** -> event features + auxiliary features
- **y** -> was event “associated” with fraud or not
- Train model on historical events, predict on real time events.

Life of Ground Truth

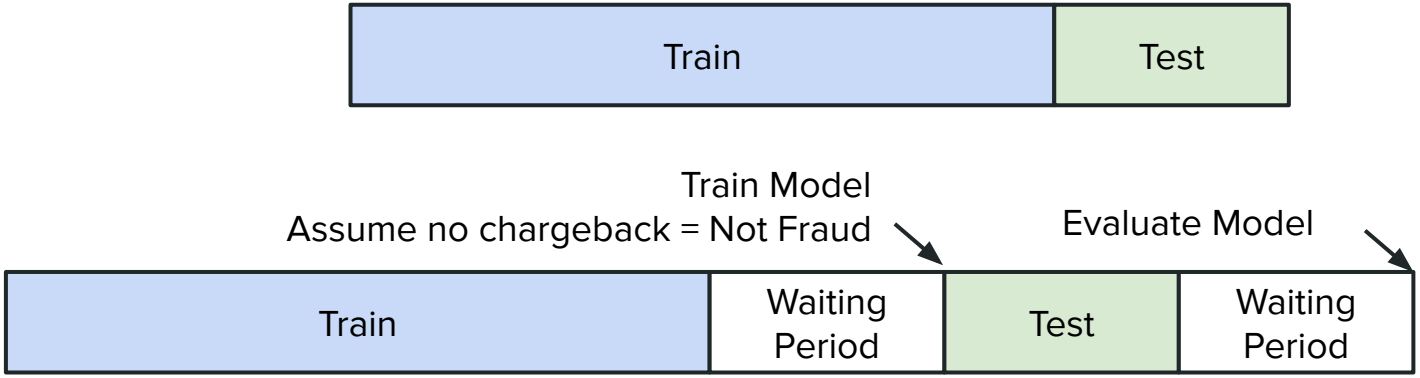


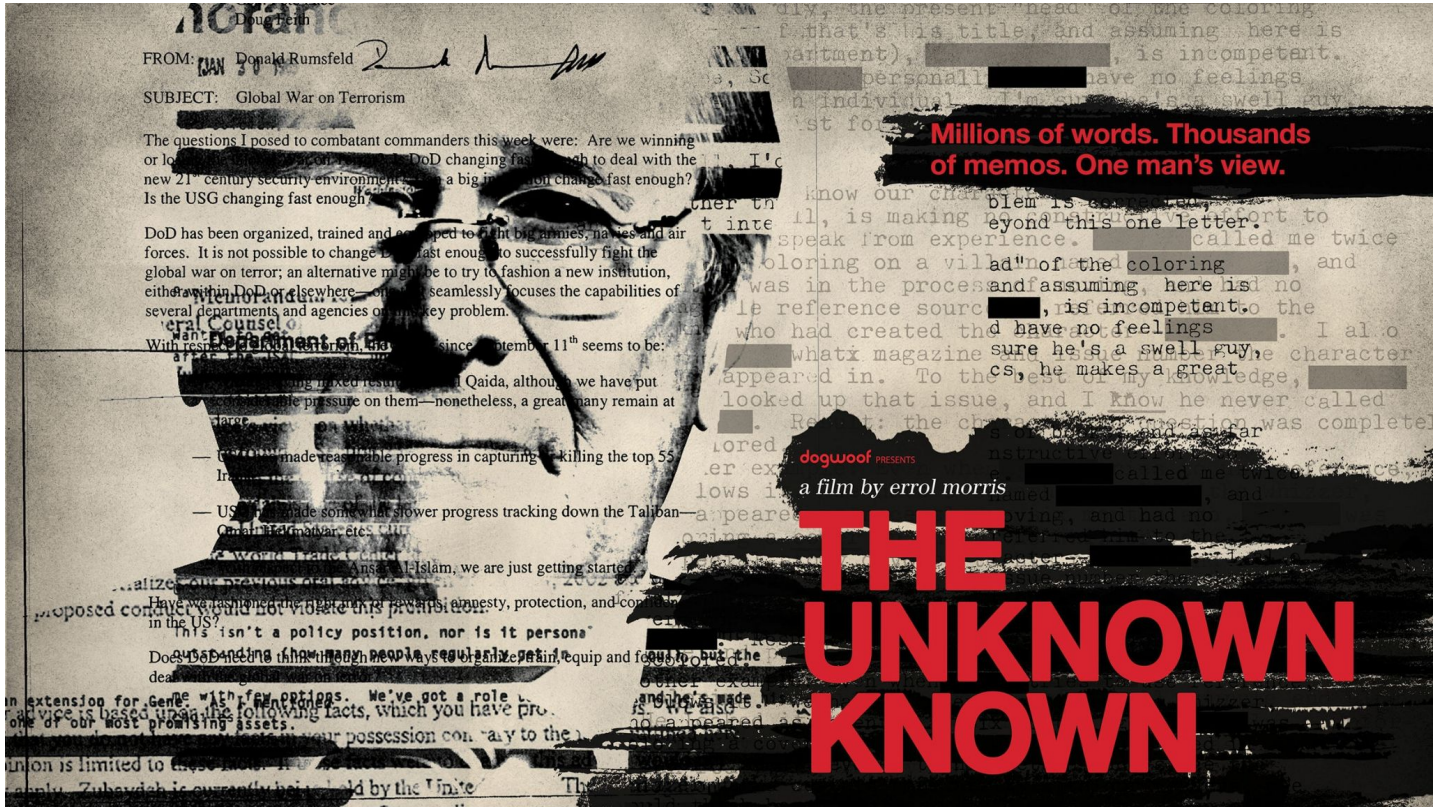
Life of Ground Truth



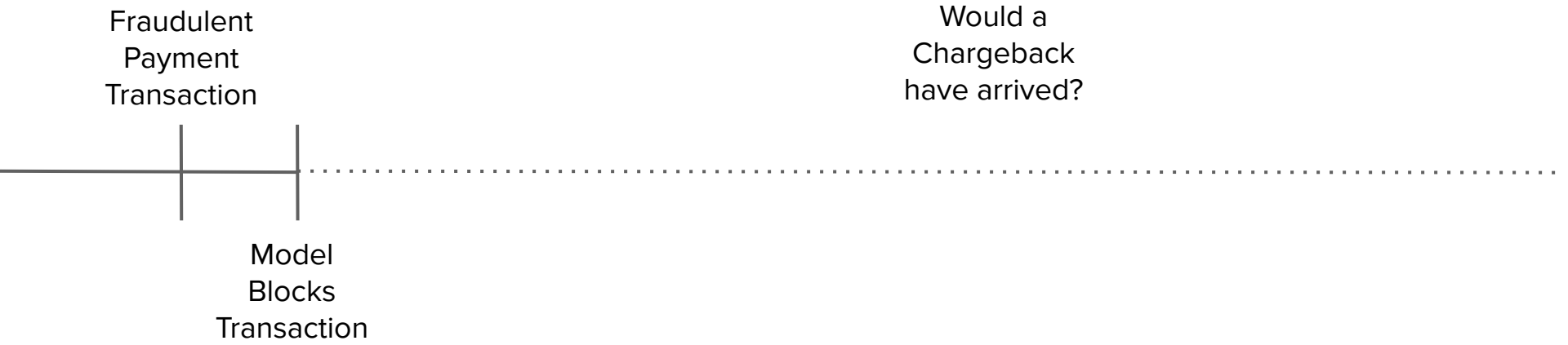
Train/Test Split

Train/Test Split

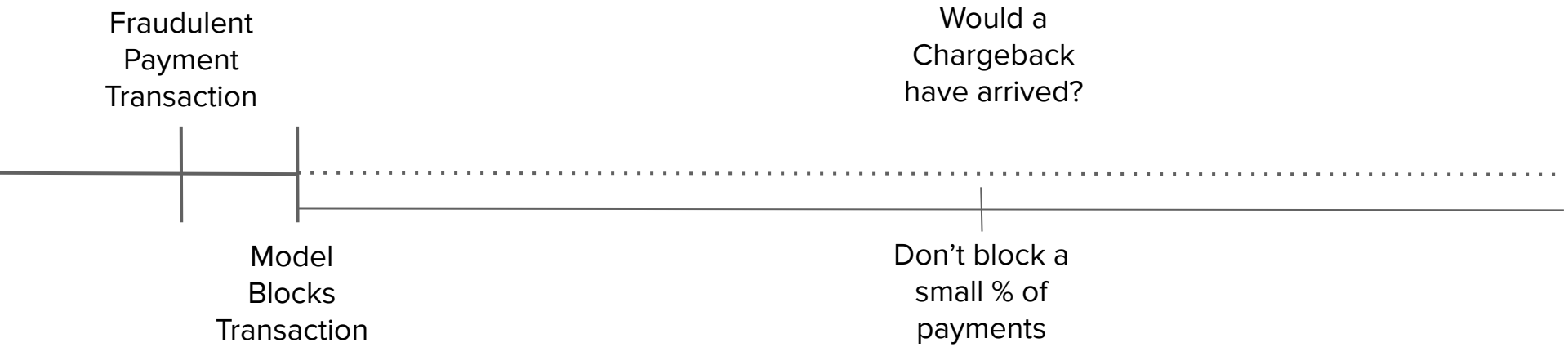



Millions of words. Thousands of memos. One man's view.
dogwoof PRESENTS
a film by errol morris
THE UNKNOWN KNOWN

Life of Ground Truth



Life of Ground Truth



Optimization

What are we optimizing for?

False Positive

- Miss out on a sale
- Churn – long term risk
- Difficult to measure!

False Negative

- Straightforward financial loss
- Could be unbounded

Capacity

- Limited capacity for human review
- Capacity has a cost

What are we optimizing for?

False Positive Precision

- Miss out on a sale
- Churn – long term risk
- Difficult to measure!

False Negative Recall

- Straightforward financial loss
- Could be unbounded

Capacity Support / Positive Prediction Rate

- Limited capacity for human review
- Capacity has a cost

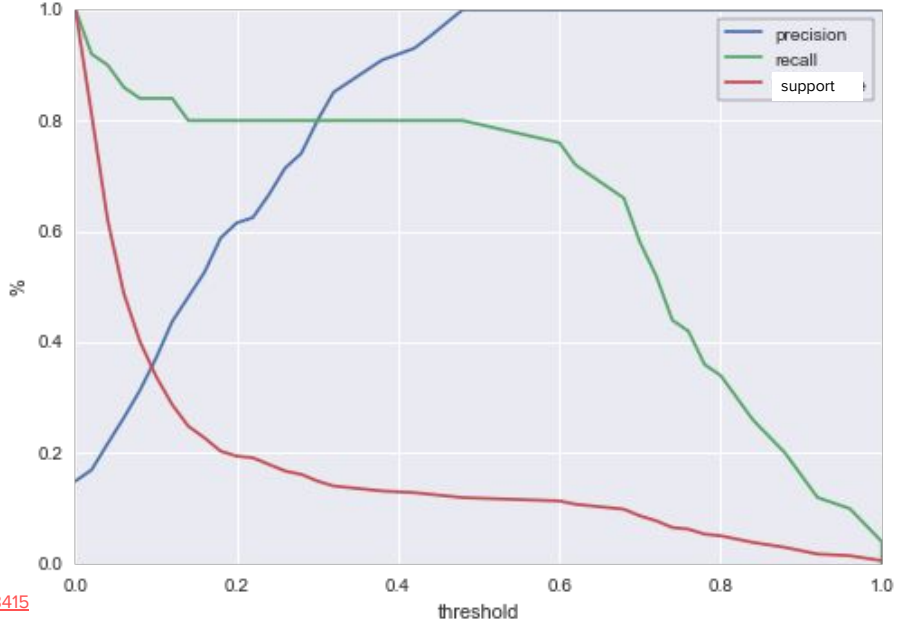
What are we optimizing for?

False Positive → Precision

- Miss out on a sale
- Churn – long term risk
- Difficult to measure!

False Negative → Recall

- Straightforward financial loss
- Could be unbounded



Capacity → Support / Positive Prediction Rate

- Limited capacity for human review
- Capacity has a cost

https://blog.insightdatascience.com/visualizing-machine-learning-thresholds-to-make-better-business-decisions-4ab07f823415

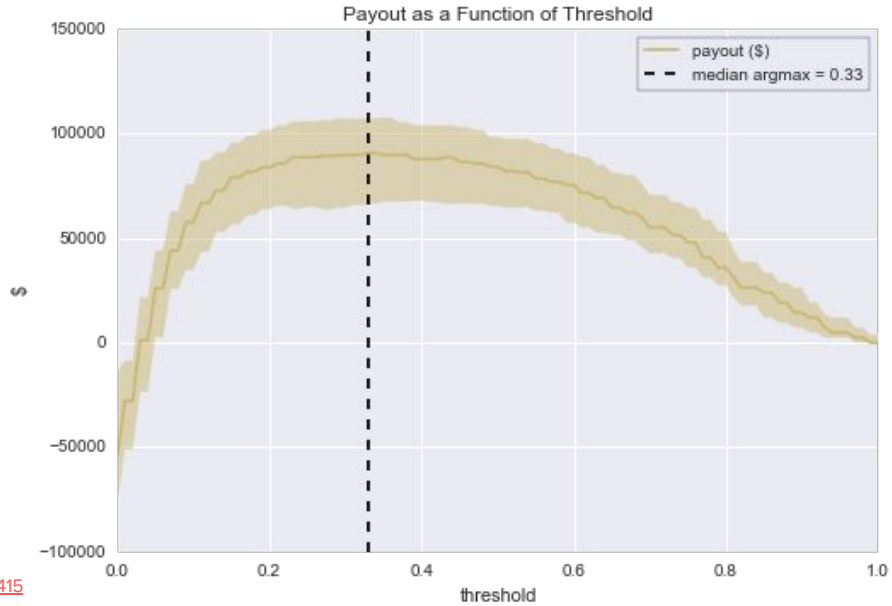
What are we optimizing for?

False Positive Precision

- Miss out on a sale
- Churn – long term risk
- Difficult to measure!

False Negative Recall

- Straightforward financial loss
- Could be unbounded



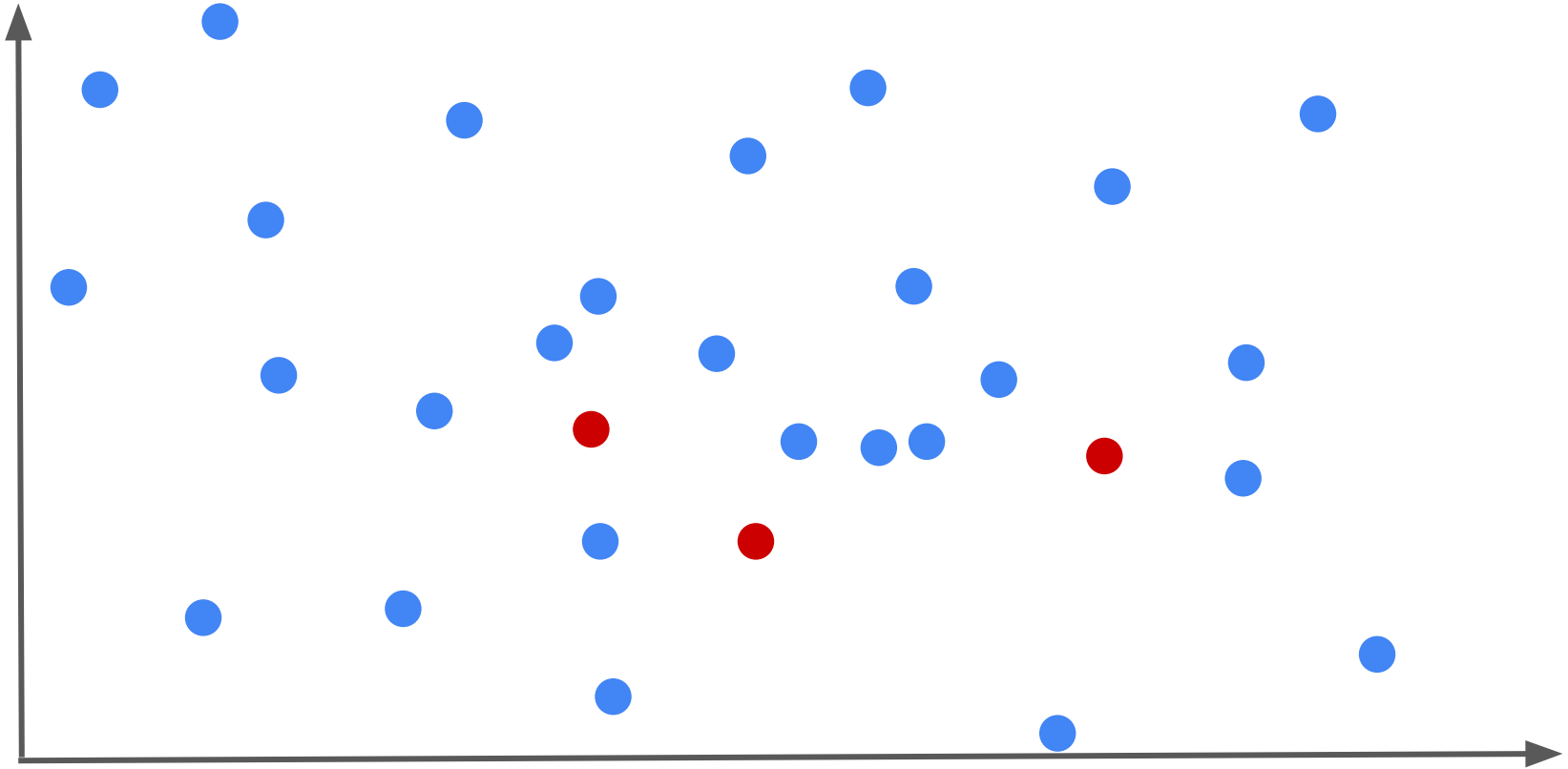
Capacity Support / Positive Prediction Rate

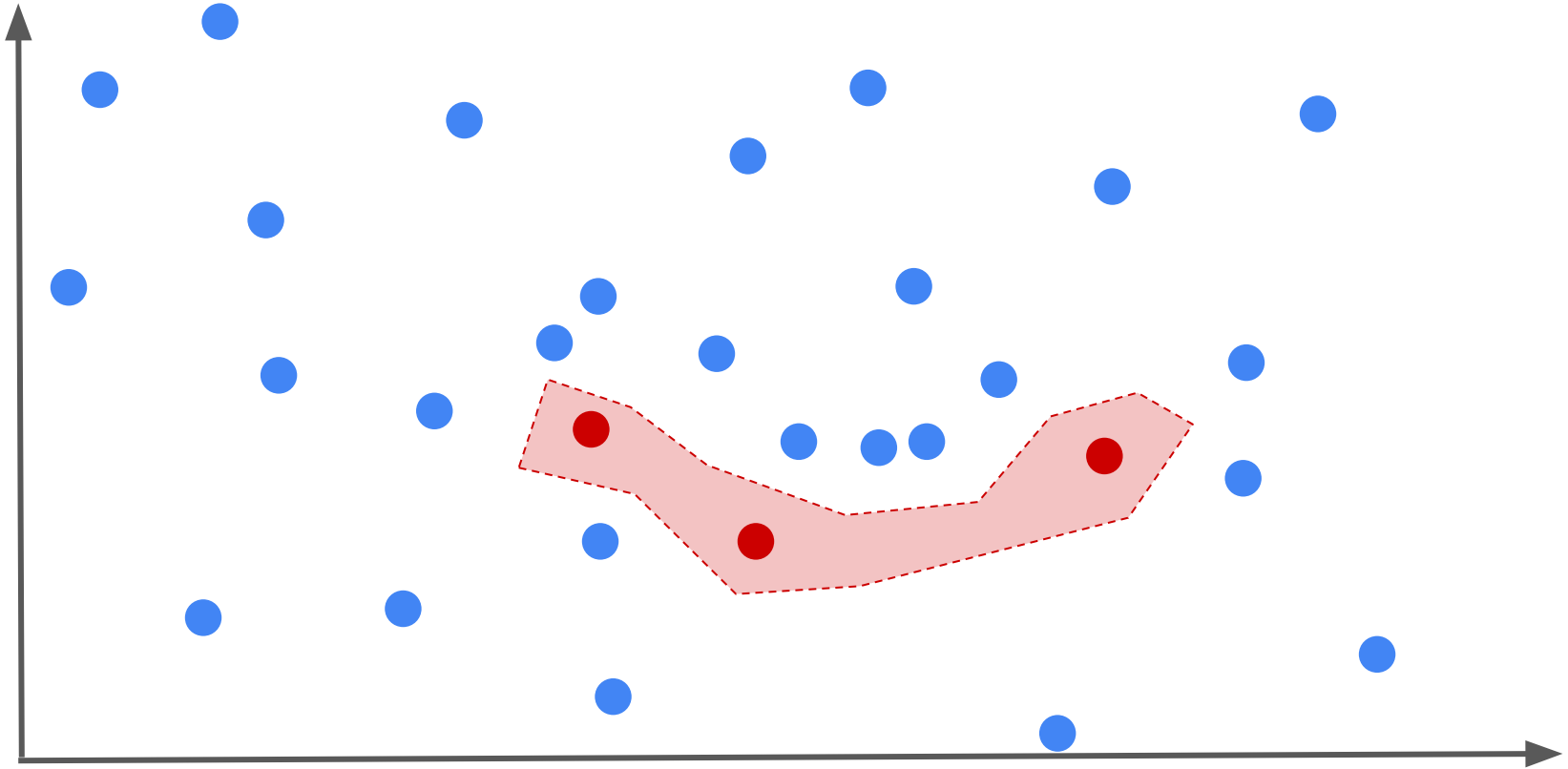
- Limited capacity for human review
- Capacity has a cost

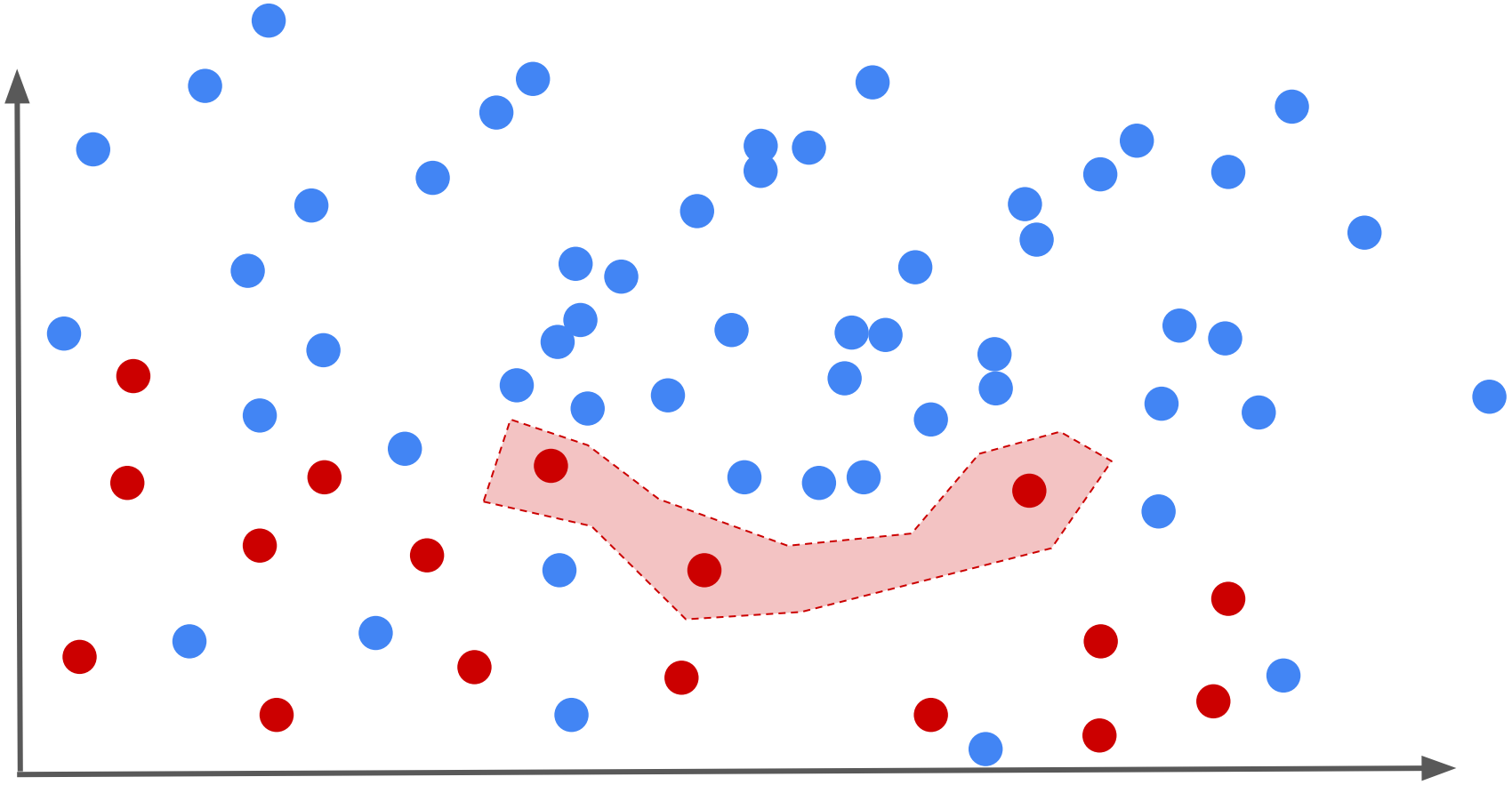
https://blog.insightdatascience.com/visualizing-machine-learning-thresholds-to-make-better-business-decisions-4ab07f823415

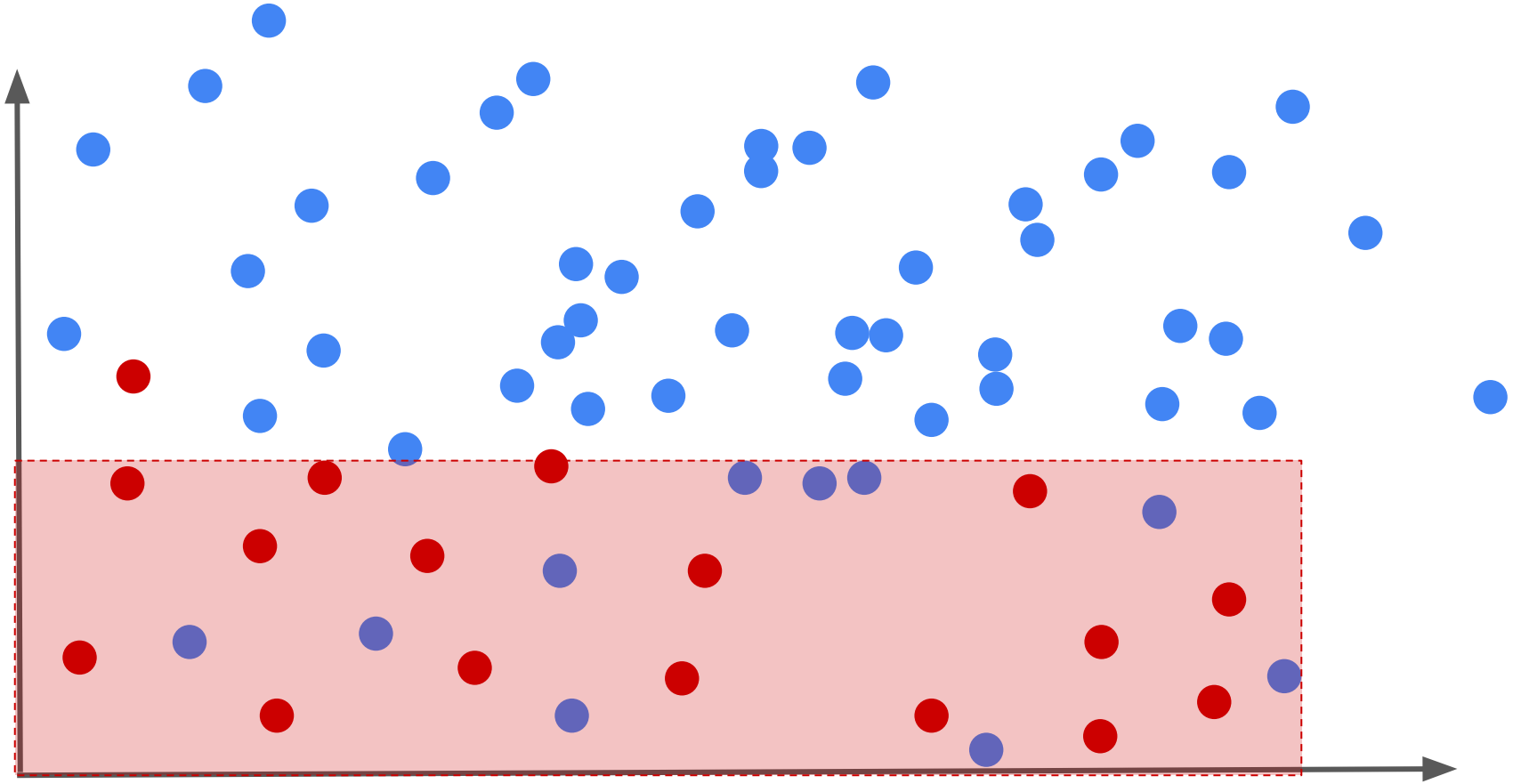
Class Imbalance

You need a lot of data



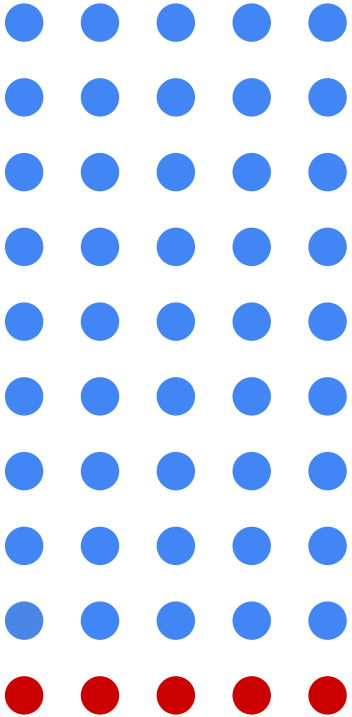


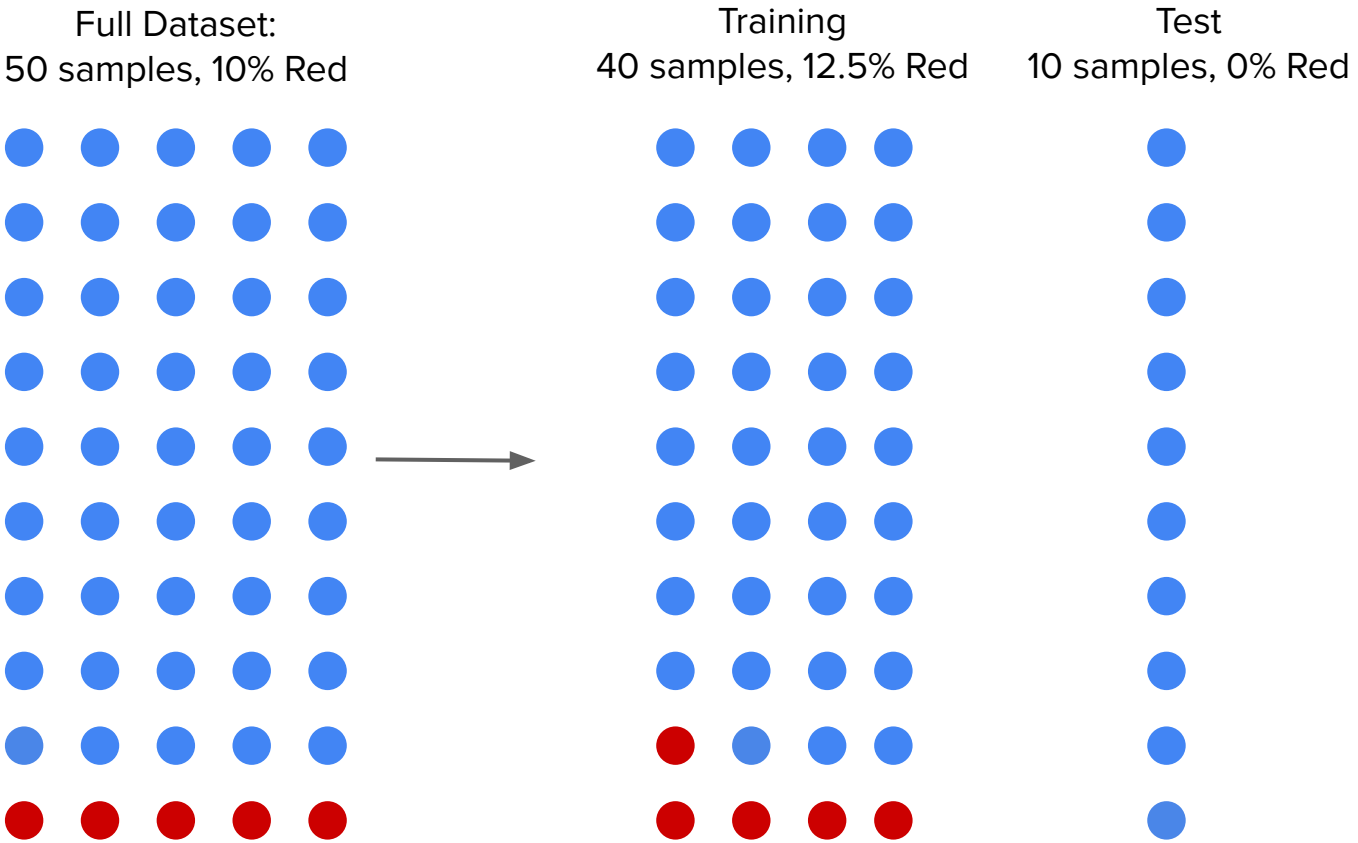


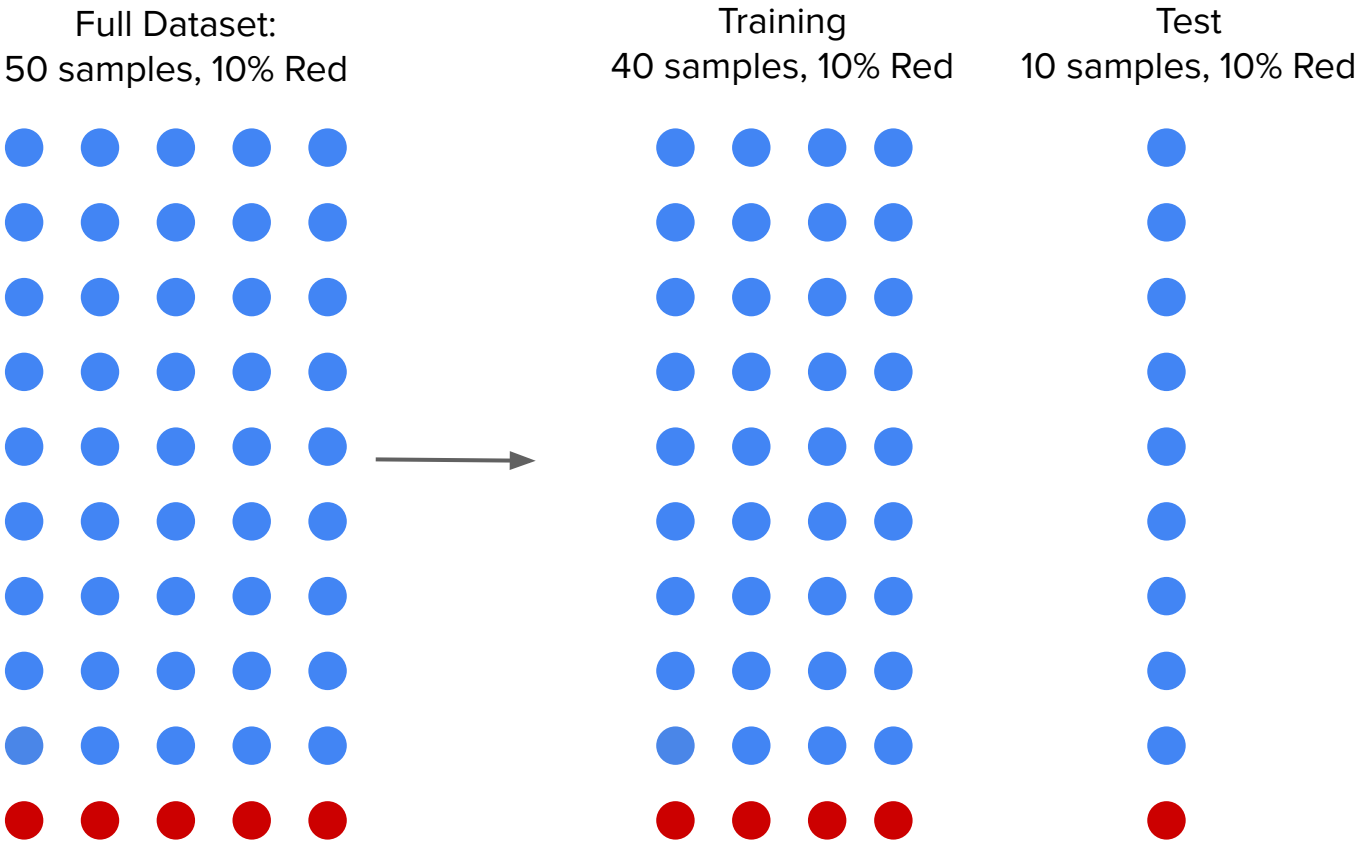


Stratified Sampling

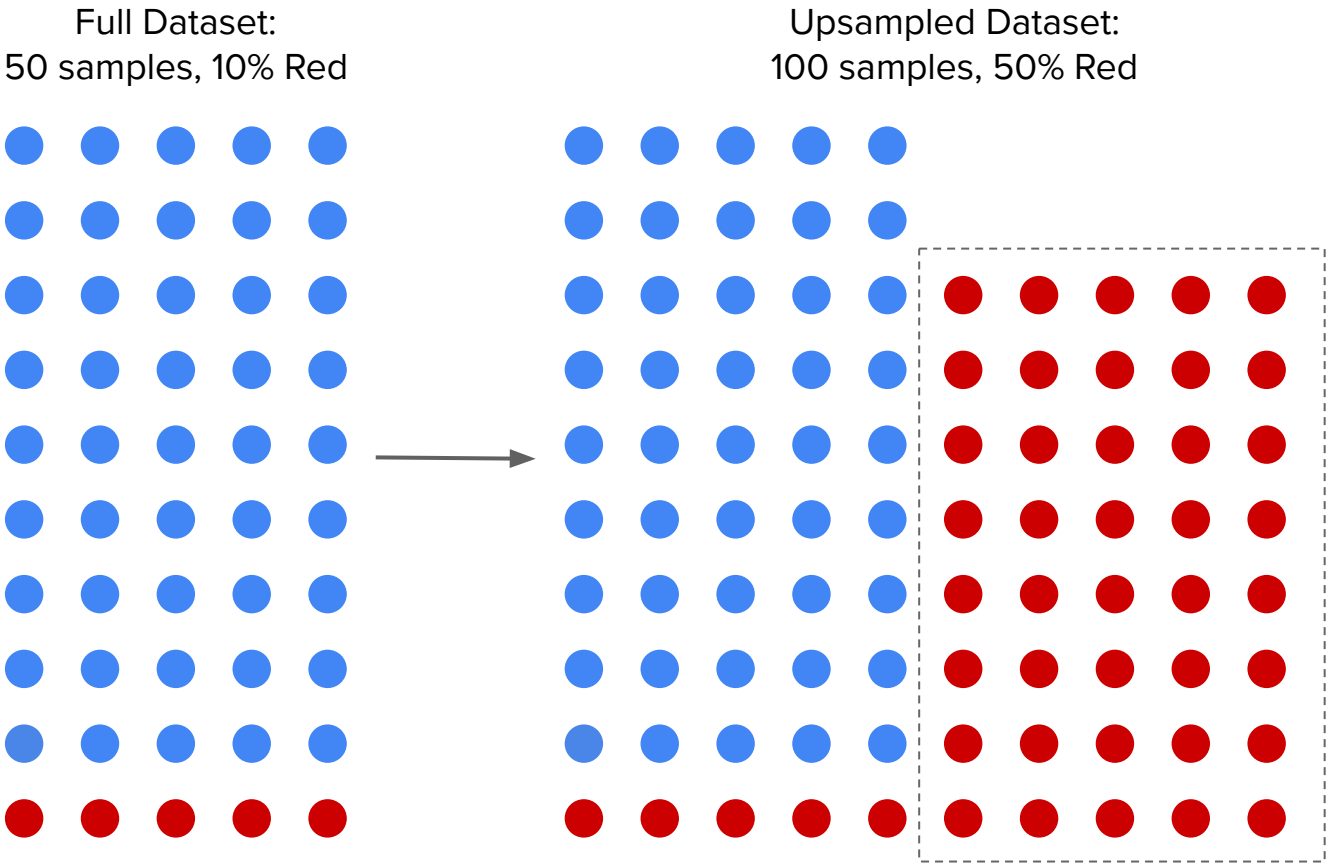
Full Dataset:
50 samples, 10% Red



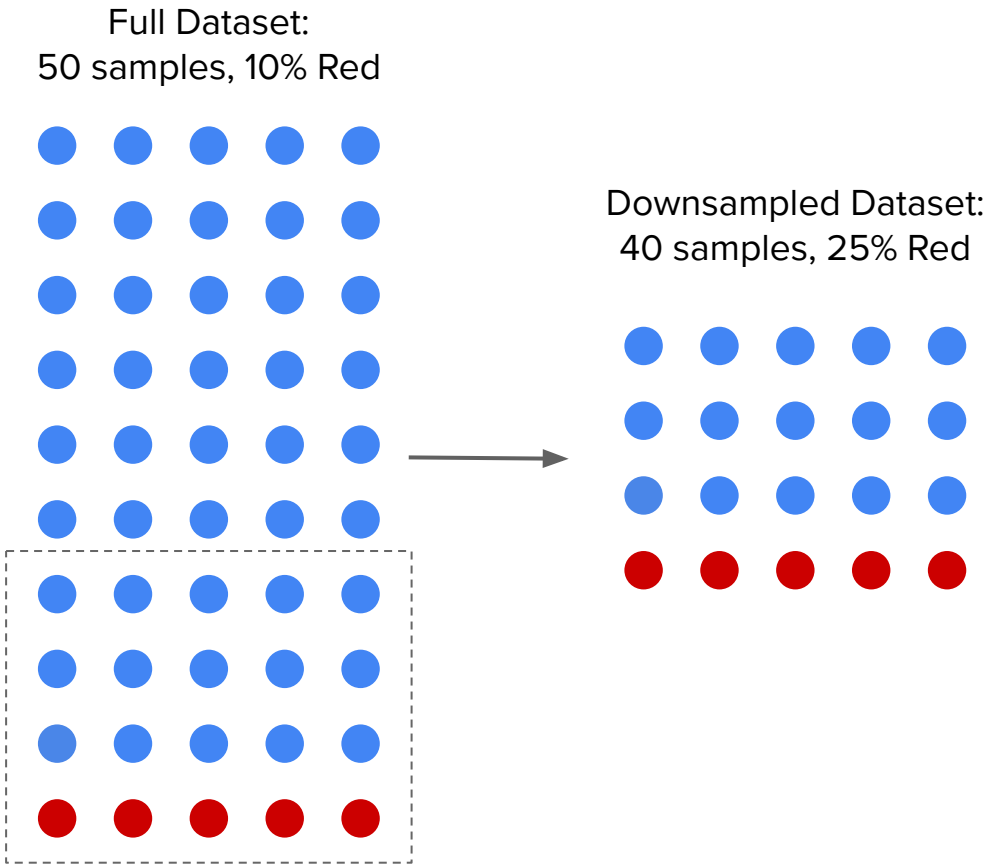

Full Dataset:
50 samples, 10% Red
Training
40 samples, 12.5% Red
Test
10 samples, 0% Red


Full Dataset:
50 samples, 10% Red
Training
40 samples, 10% Red
Test
10 samples, 10% Red

Upsampling

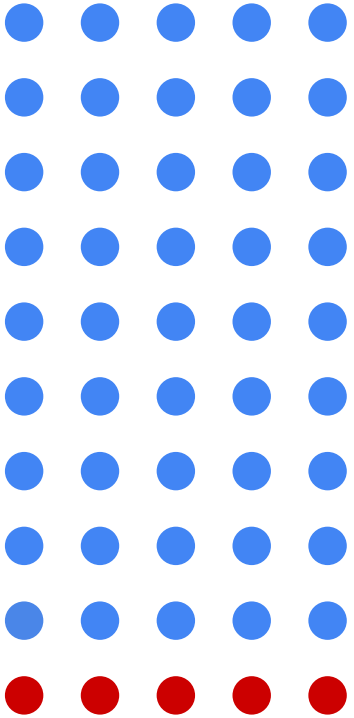

Full Dataset:
50 samples, 10% Red
Upsampled Dataset:
100 samples, 50% Red

Downsampling

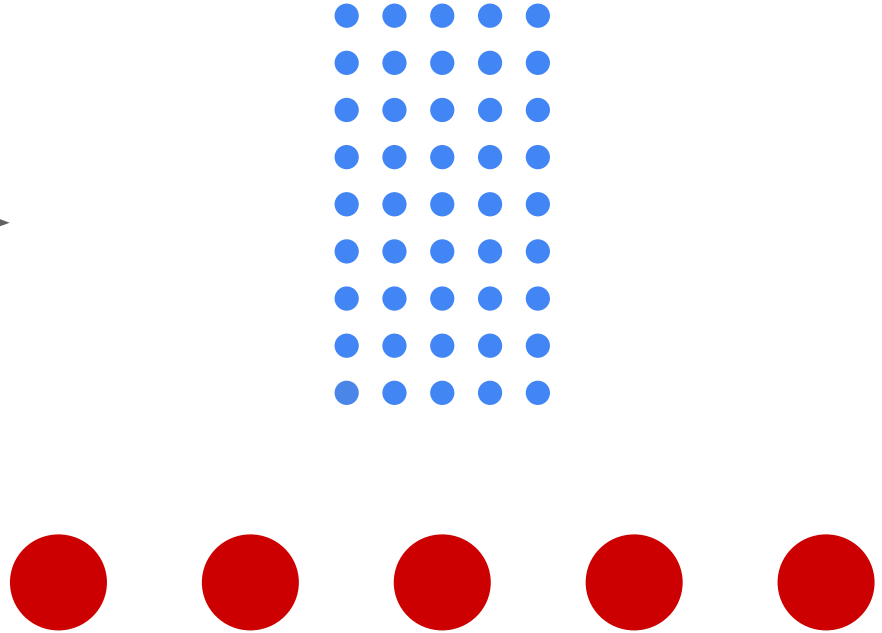

Full Dataset:
50 samples, 10% Red
Downsampled Dataset:
40 samples, 25% Red

Class Weighting

Full Dataset:
50 samples, 10% Red



Class Weighted Dataset



Metrics

Imbalanced Metrics

- Accuracy is not good!
- Sampling -> weighted metrics
- Uncertainty