Operationalizing machine learning models for DNS security

Thymen Wabeke, Thijs van den Hout
TMA22 PhD School

27 June 2022

SIDN is the operator of the .nl ccTLD

Registration of domain names

6.2M .nl-domains

Publish domain names via DNS

2.5B DNS queries/day

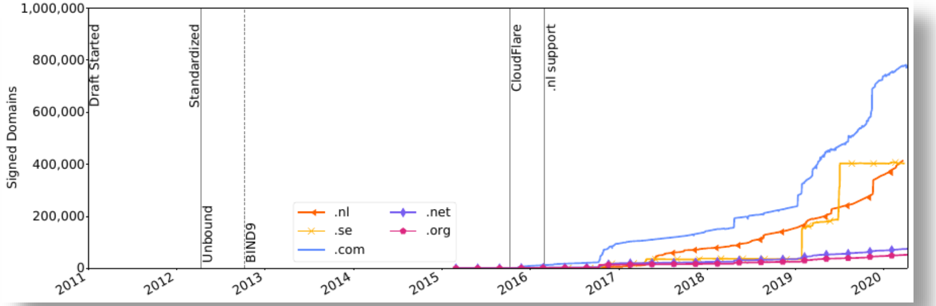
SIDN Labs = research team

- Goal: increase trustworthiness of our society's internet infrastructure, for .nl and the Netherlands in particular.

- Strategies:
  - Applied technical research (measurements, design, prototyping, evaluation)
  - Make results publicly available and useful for various target groups
  - Work with universities, infrastructure operators, and other labs

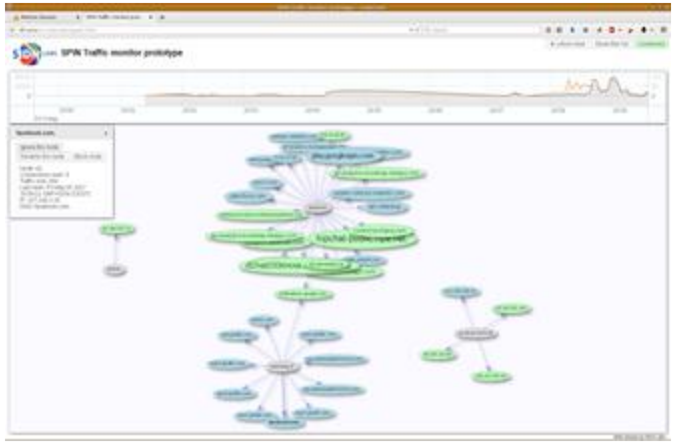
- Three research areas: network security (DNS, NTP, BGP), domain name & IoT security, secure future internet infrastructures

Example projects

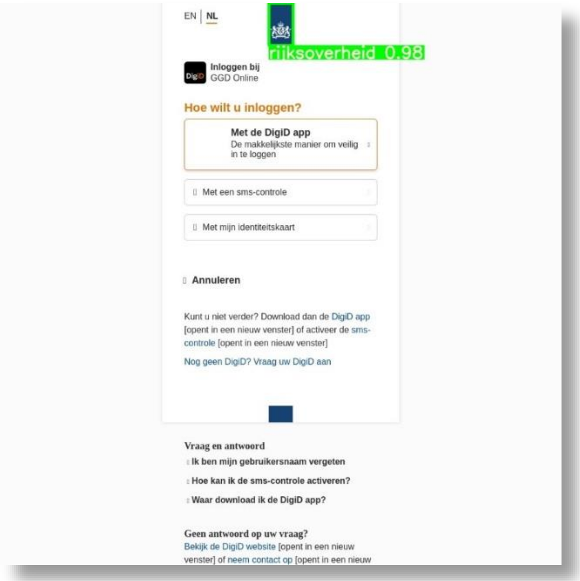


Measuring the deployment of newly standardized DNSSEC algorithms [3]

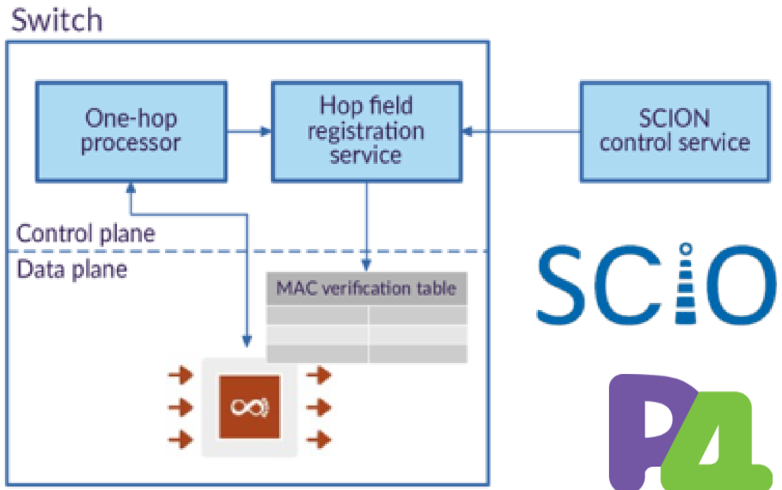
Provide well-managed and secure time services [4]



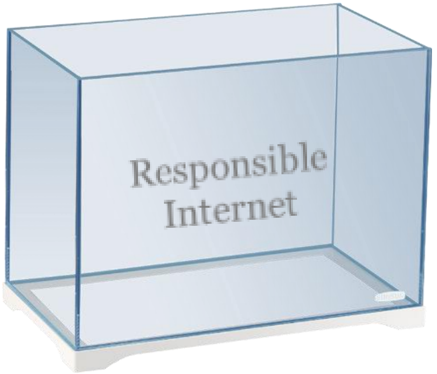
Making the IoT more secure and transparent and measure its evolution [5]



Logo detection technology to identify malicious .nl websites [6]



Experimenting with secure future networks and programmable networks [7][8]



€1.9M

Developing a new Internet security and autonomy paradigm [9]

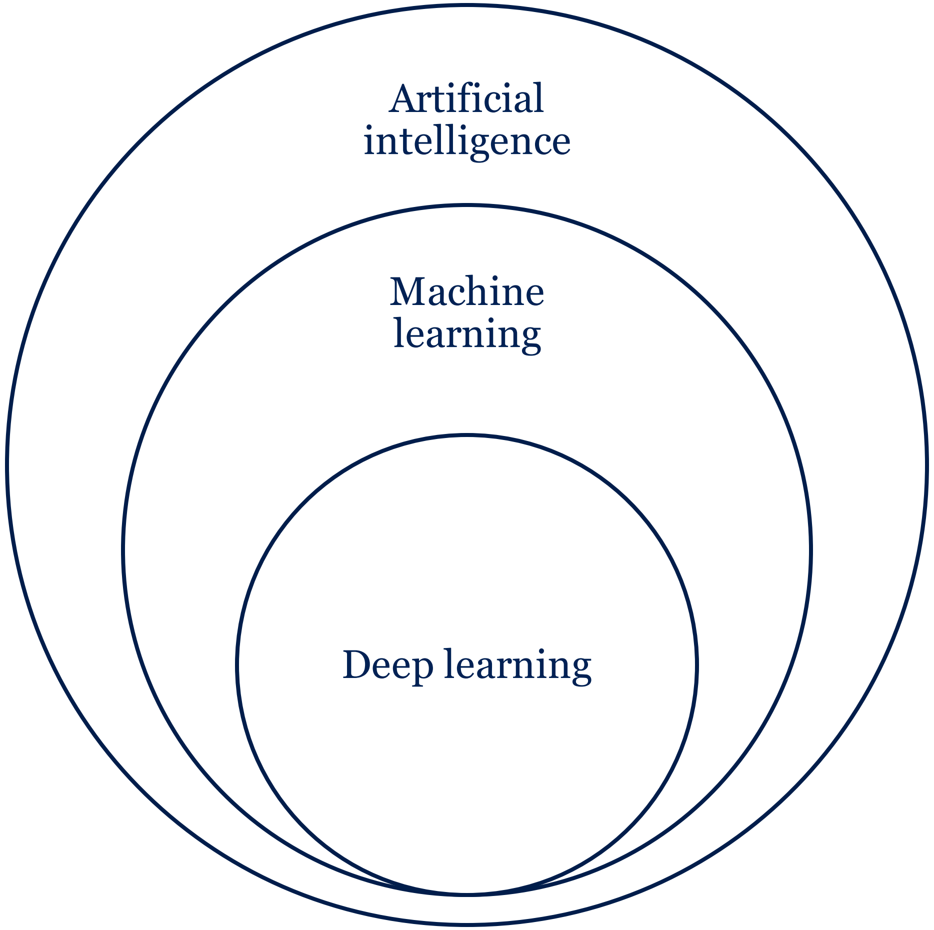
Today's agenda

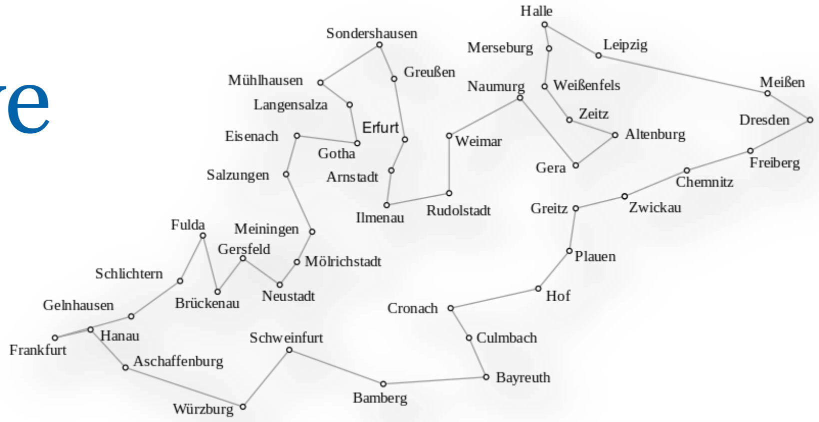
1. Successful ML applications [30 min]
2. ML with an operational mindset [20 min]

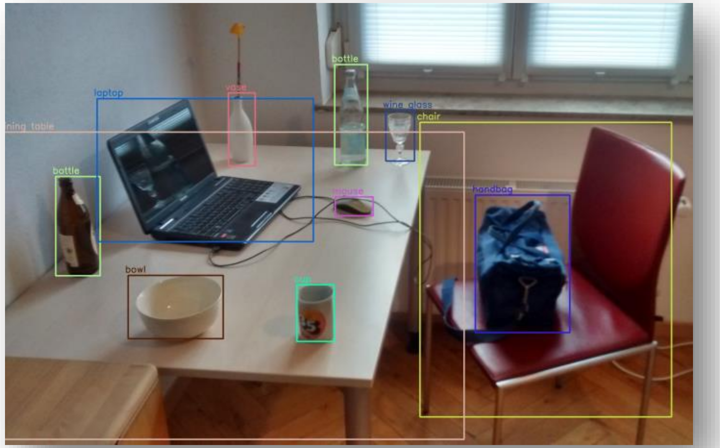
Break

3. Train, evaluate and tune a fraud detection classifier [40 min]
4. Improve classifier using active learning [40 min]

Machine learning in perspective



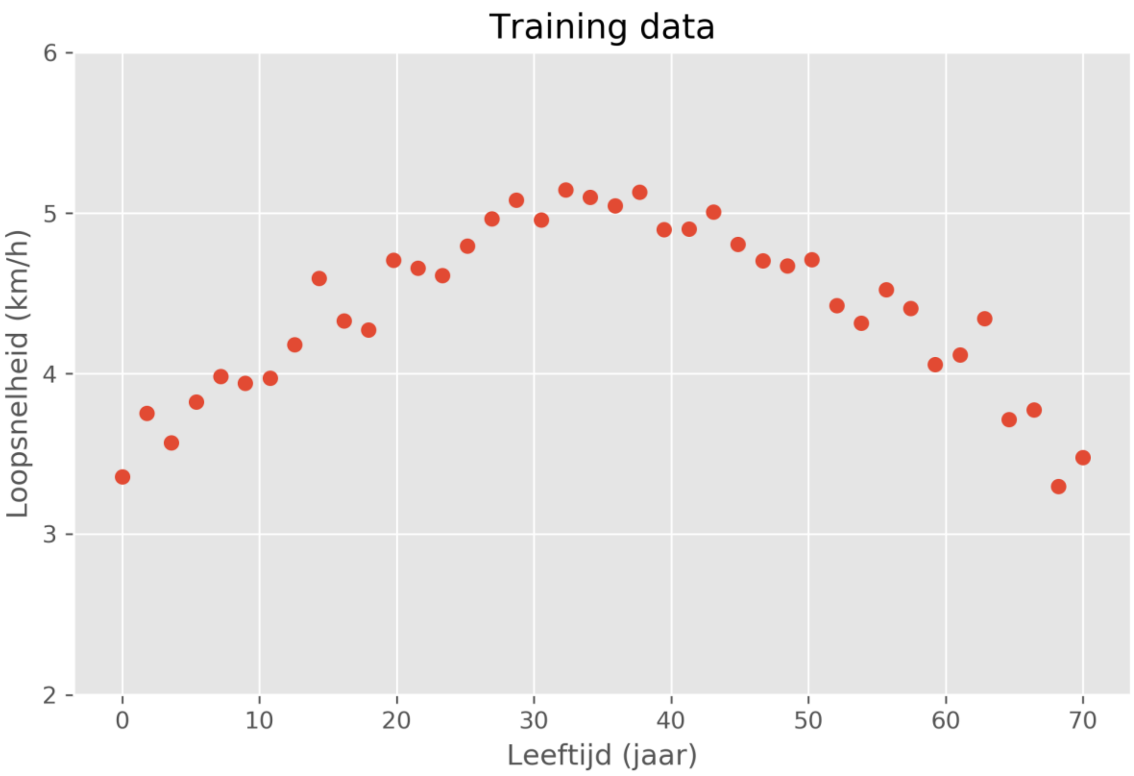




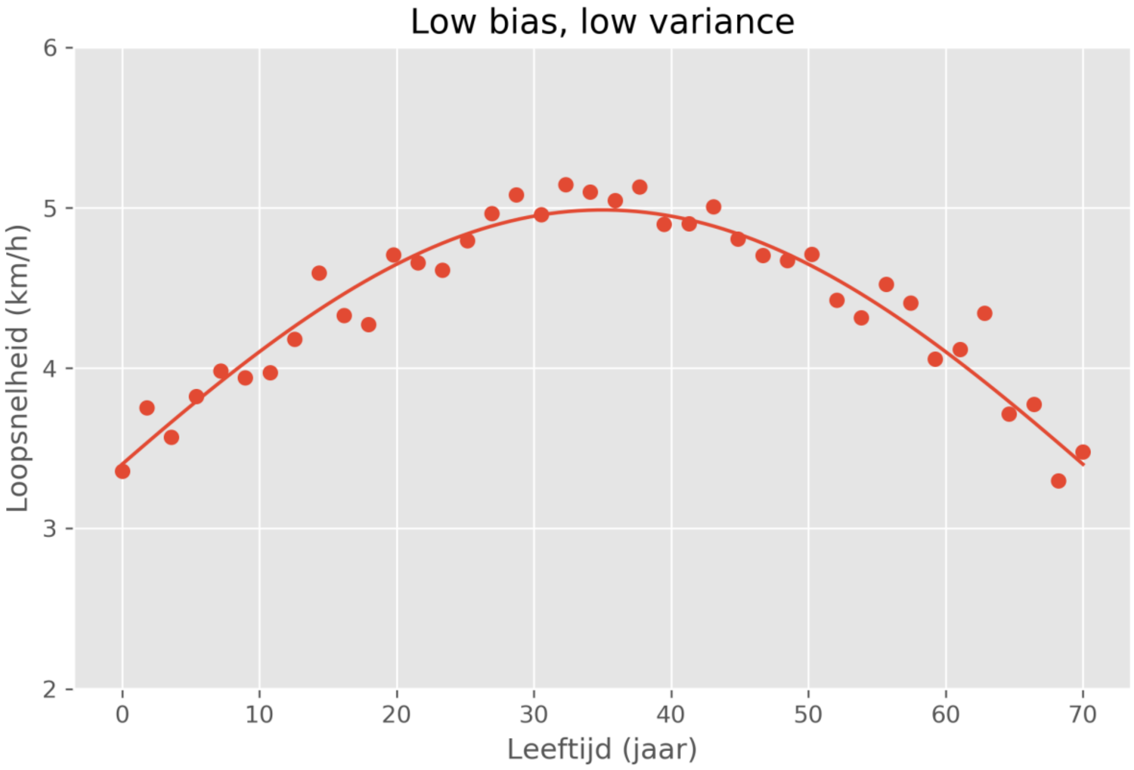
Learning paradigm

Train → **Test** → **Infer**

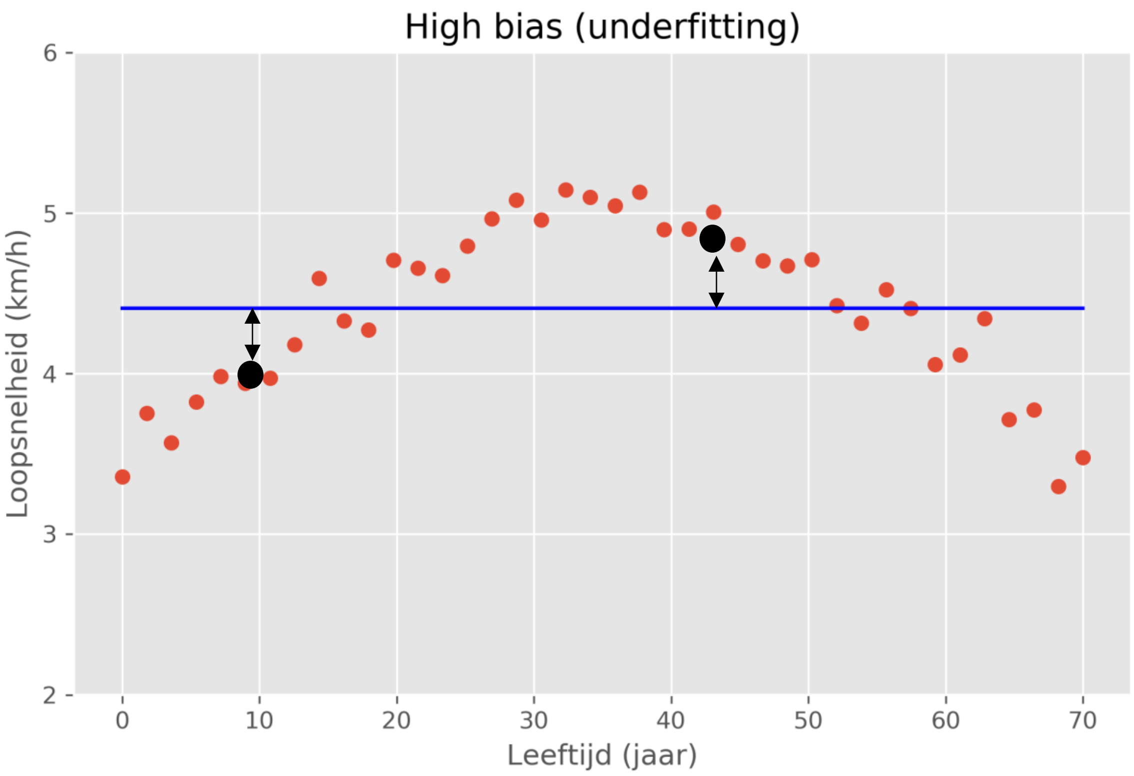
Training a regression model



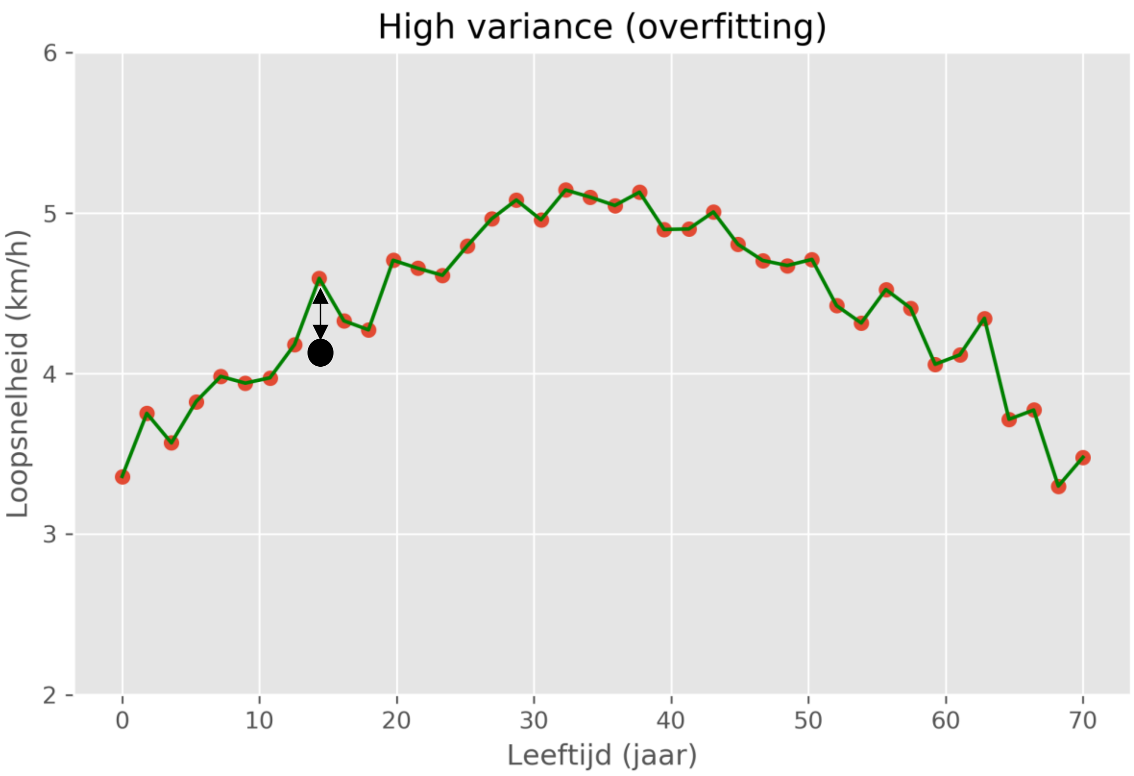
Training a regression model



Training a regression model



Training a regression model



This only works when...

- Data and ground truth labels are available
- Labels are well defined
- Data is representative

Research agenda

- Apply ML to increase security of the Internet and DNS

- Approach: explore and integrate promising algorithms, papers and tools
  - Innovating *with* ML, not innovation *of* ML

- Target group: DNS actors (registries, registrars and DNS operators)

See also: https://www.sidnlabs.nl/en/news-and-blogs/using-new-machine-learning-applications-to-boost-internet-security

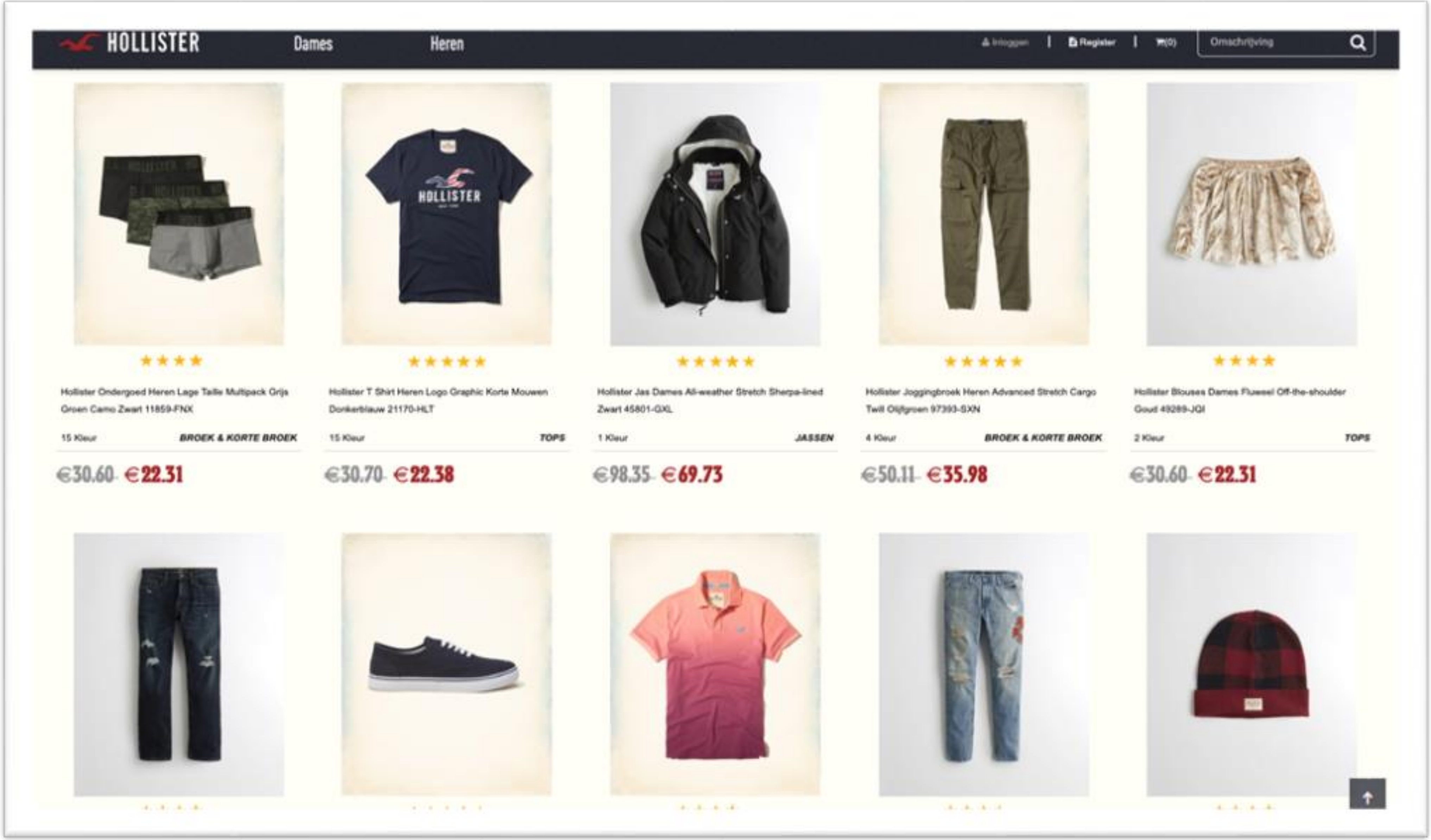
SIDN LABS

Two successful machine learning projects

FaDe

LogoMotive

nederlandwebshop.nl



SIDN's interest

- Consumer losses
- Trust in Internet may decrease

Perfect vantage point:

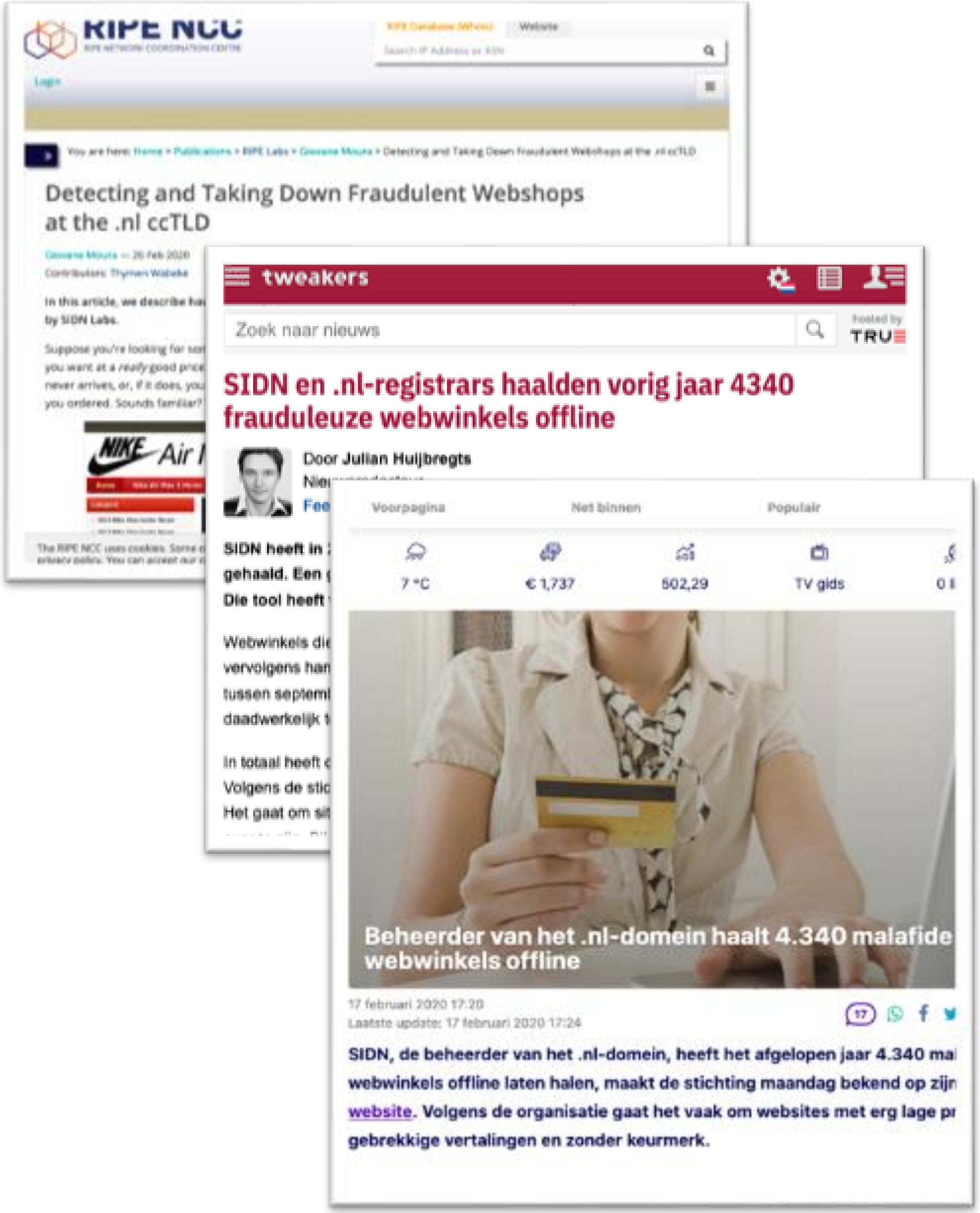
- List of *all* .nl-domains
- Passive and active measurements

SIDN LABS

Main results

- Detected thousands since 2016
- Protected users from being scammed

- PAM2020 paper:
  - BrandCounter (2018 Q1-2)
  - FaDe (2019 Q1)



Published at PAM2020: https://doi.org/10.1007/978-3-030-44081-7_10
Video at RIPE80: https://ripe80.ripe.net/archives/video/322/

- Support Vector Machine
- Optimized using grid search

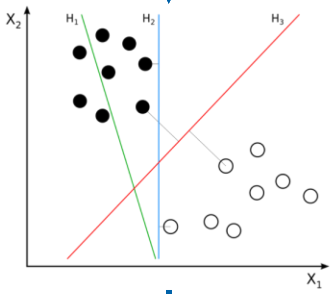
Dataset

- 231 counterfeit
- 229 legitimate

Features

- 6 registration
- 3 infrastructure

Training samples → **Train model**



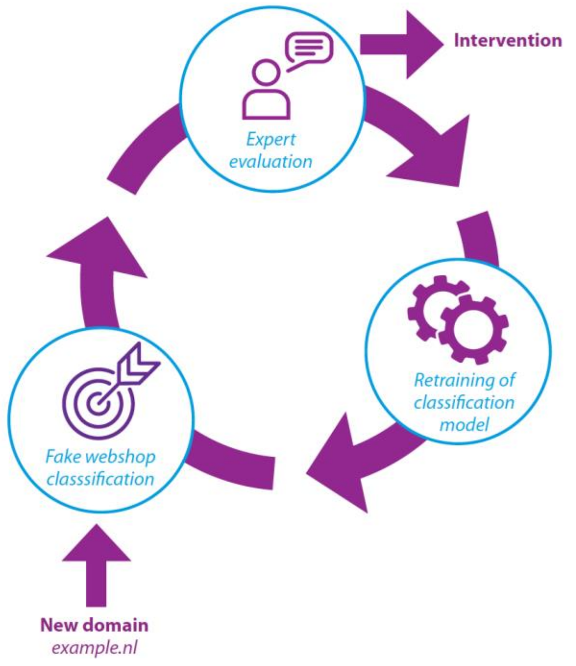
Testing samples → **Apply model**

Samples	Precision	Recall
Train (cross-validation)	0.98	0.97
Test	1.0	1.0

SIDN LABS

Lessons learned

- Registrar and ICS collaboration was key
- Detectors are simple yet effective
  - Registries have perfect vantage point
  - Suggests little pressure
- It’s an ever-going whack-a-mole game

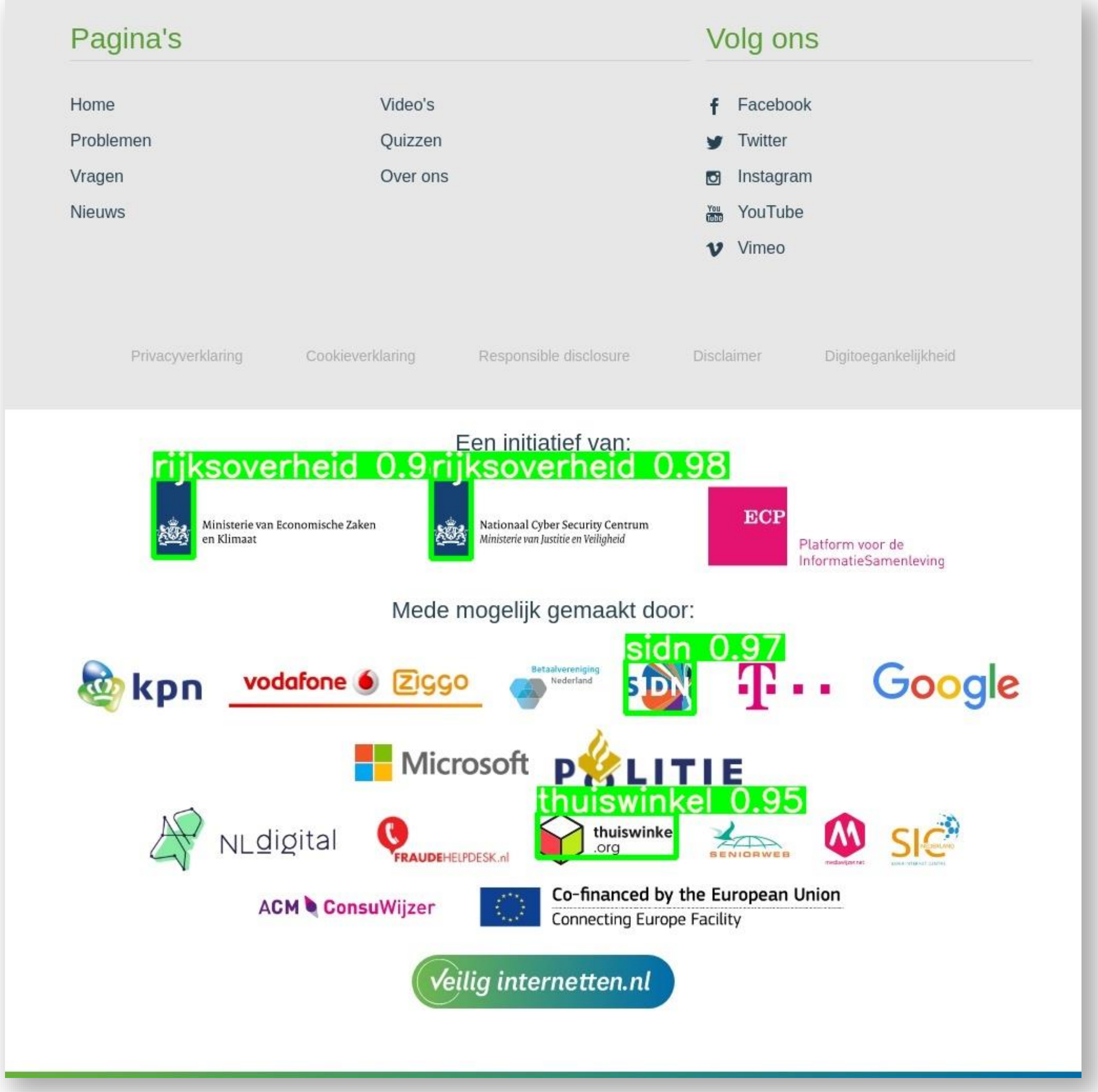


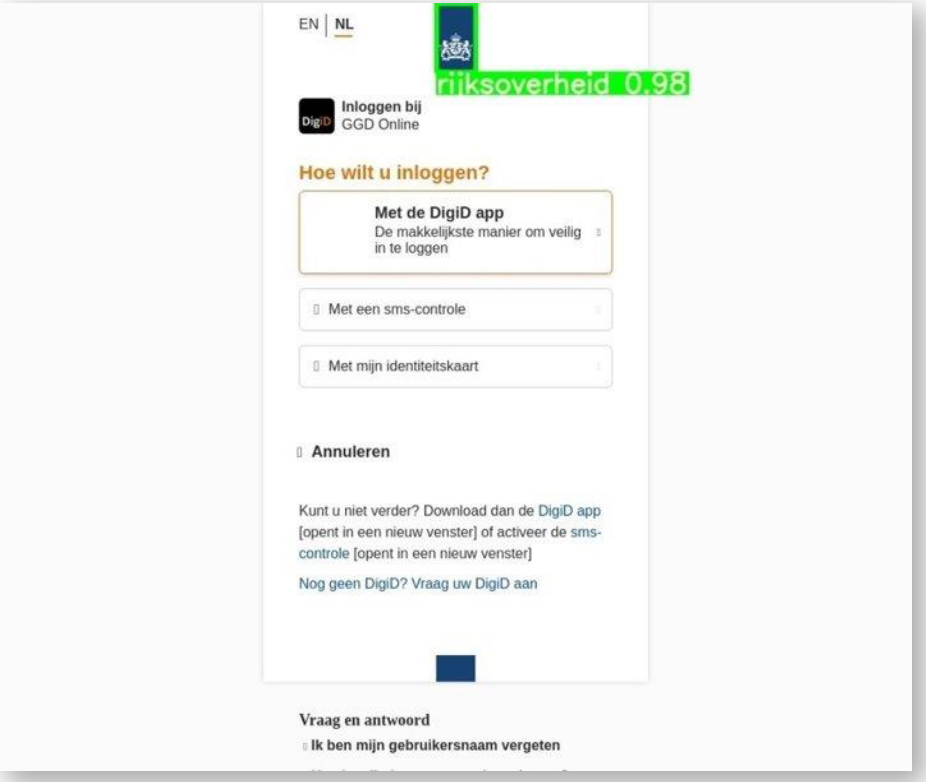
Year	Taken down
2018	~12,000
2019	4,340
2020	481

Number of counterfeit webshops taken down

SIDN LABS

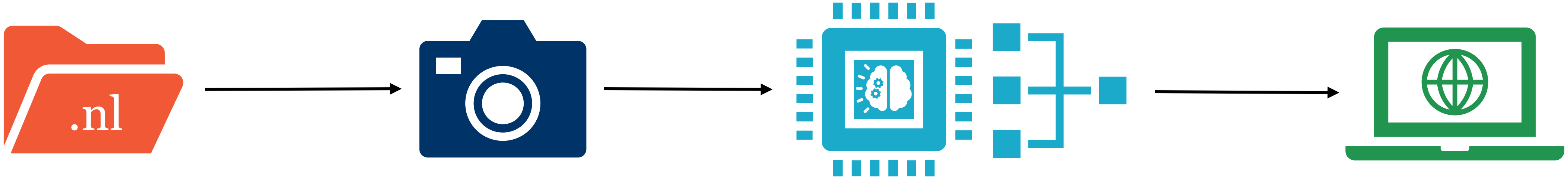
LogoMotive: finding malicious .nl-domains with logo detection





SIDN LABS

How does LogoMotive work?



List of .nl
Domain names

Automatically visit
and screenshot
websites

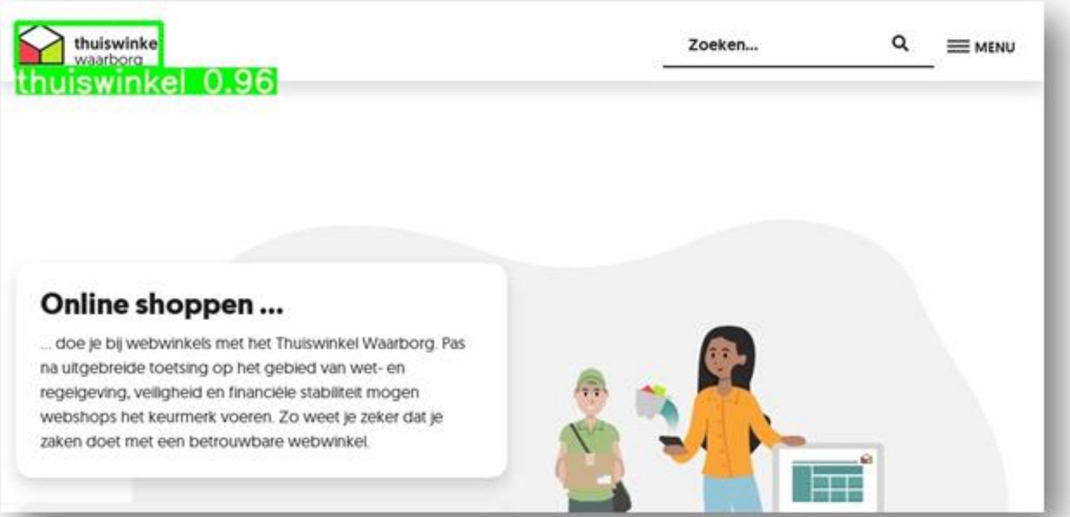
Apply logo detection
to the screenshots

Upload results to
online dashboard

Can logo detection contribute to a safe .nl-zone?

Case study with Dutch national government

Found: Phishing, suspicious redirects, security threats



Case study with Dutch webshop trustmark (Thuiswinkel.org)

Found: Trustmark abuse, improved domain portfolio

More info & paper: logomotive.sidnlabs.nl

Machine learning with an operational mindset

Part 2/4

Use case: detect suspicious registrations

- 22%-62% of abusive domains were registered with malicious intents
  - Phishing, malware, DGAs

- Verifying new registrations could prevent malicious registrations
  - But: +/- 2500 registrations per day
  - But: reviewing a registration takes 5-20 minutes
  - But: only 3 (0.11%) reported at Netcraft within 30 days

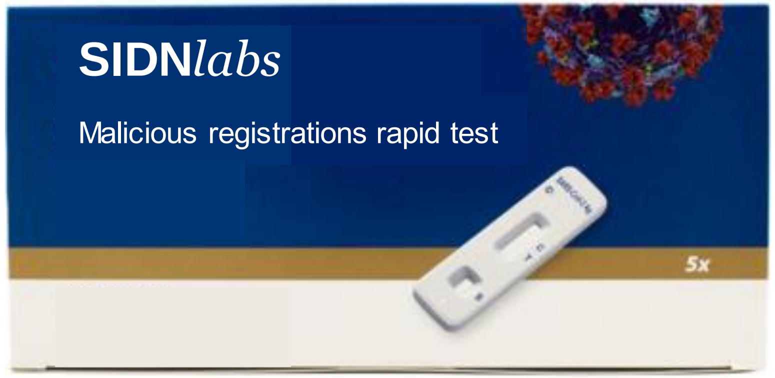
SIDN LABS

Goal: identify registrations that should be reviewed

- Classify whether a registration is benign or suspicious
  - Only data available at registration

- Support will manually review suspicious registrations
  - No algorithmic decision making

- Prevent scams
  - Not verifying clearly benign registrations



Research vs. operational environment

- Project is suitable for:
  - Research project at a university (outcome = paper)
  - Operational project within an organization (outcome = deployed classifier)

- How will developing the classifier differ between these 2 environments?

Research vs. operation: identify differences

Go to www.menti.com and enter 3393 6819

Train, evaluate & tune a fraud detection classifier

Part 3/4

We start at 15:10

SIDN LABS

Characteristics of classification problem

	RegCheck	TransactCheck
Row	New domain name registrations	Credit card transactions
Number of rows	~ 900k in 2021	~ 286k for a year
Class labels	Class 0: Not reported Class 1: Reported within 28 days	Class 0: Legitimate Class 1: Fraudulent
Goal	Detect malicious registrations	Detect fraudulent transactions
Abuse ratio	~ 0.11%	~ 0.17 %
Labelling costs	Strong labels expensive	Strong labels expensive
Input	Domain name, registrar, creation date, name servers, name and address details of registrant.	Transaction amount, 28 unnamed features which are components generated by a PCA
Sensitivity	Many PIDs	No PIDs due to PCA

SIDN LABS

Characteristics of classification problem

	RegCheck	TransactCheck
Row	New domain name registrations	Credit card transactions
Number of rows	~ 900k in 2021	~ 286k for a year
Class labels	Class 0: Not reported Class 1: Reported within 28 days	Class 0: Legitimate Class 1: Fraudulent
Goal	Detect malicious registrations	Detect fraudulent transactions
Abuse ratio	~ 0.11%	~ 0.17 %
Labelling costs	Strong labels expensive	Strong labels expensive
Input	Domain name, registrar, creation date, name servers, name and address details of registrant.	Transaction amount, 28 unnamed features which are components generated by a PCA
Sensitivity	Many PIDs	No PIDs due to PCA

Assignment 1: Develop a TransactCheck model

- Explore dataset
- Train 2 or more `scikit-learn` models using balanced dataset of 2 weeks
  - At least 1 interpretable model
- Tune and test models using holdout data
  - Precision vs. recall tradeoff
  - Choose a threshold

Instructions

1. Find a coding partner

2. Browse to https://colab.research.google.com and sign-in with a Google Account

3. New to Google Colab and/or Jupyter Notebook?
   Browse to https://colab.research.google.com/notebooks/intro.ipynb

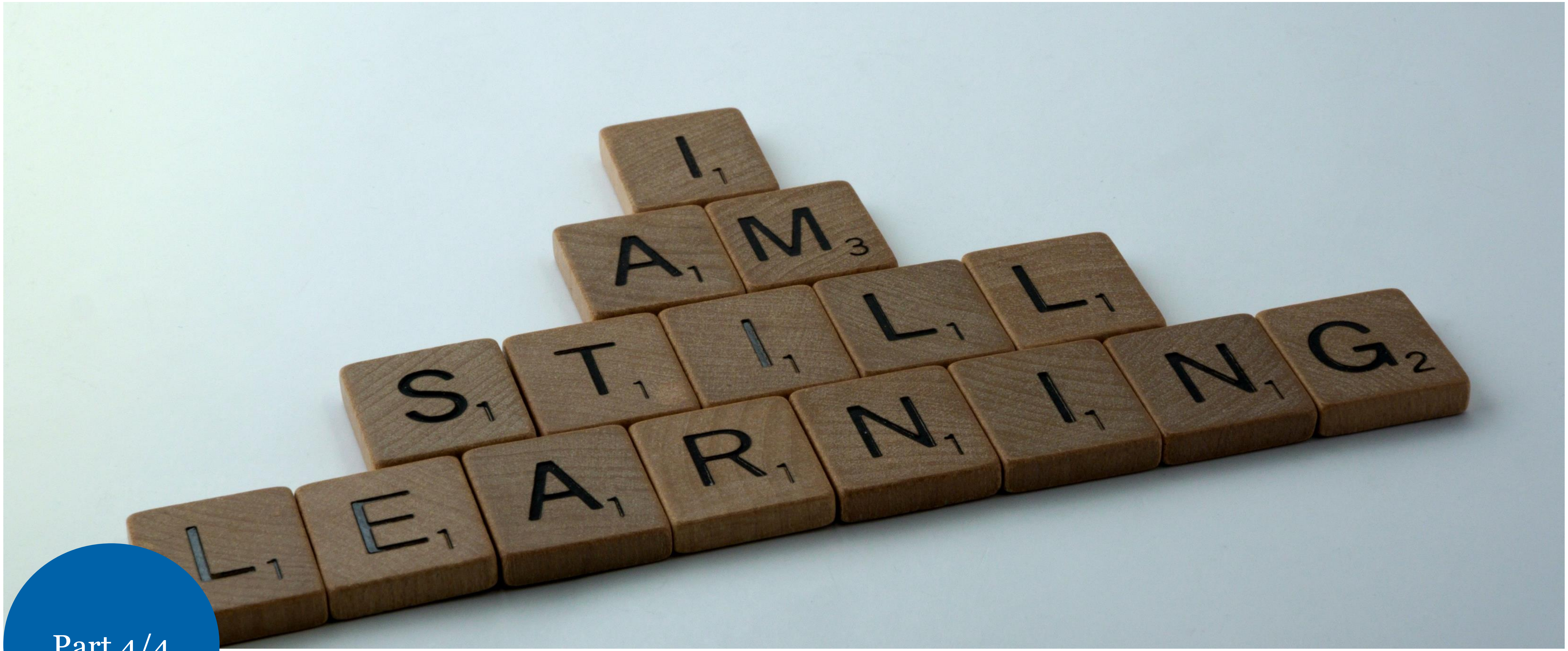
4. Ready for the real deal?
   Browse to **github.com/SIDN/tma22_ml** and click on the Assignment1 link in the README

SIDN LABS

Results assignment 1

Go to www.menti.com and enter 3393 6819

Improve classifier using active learning

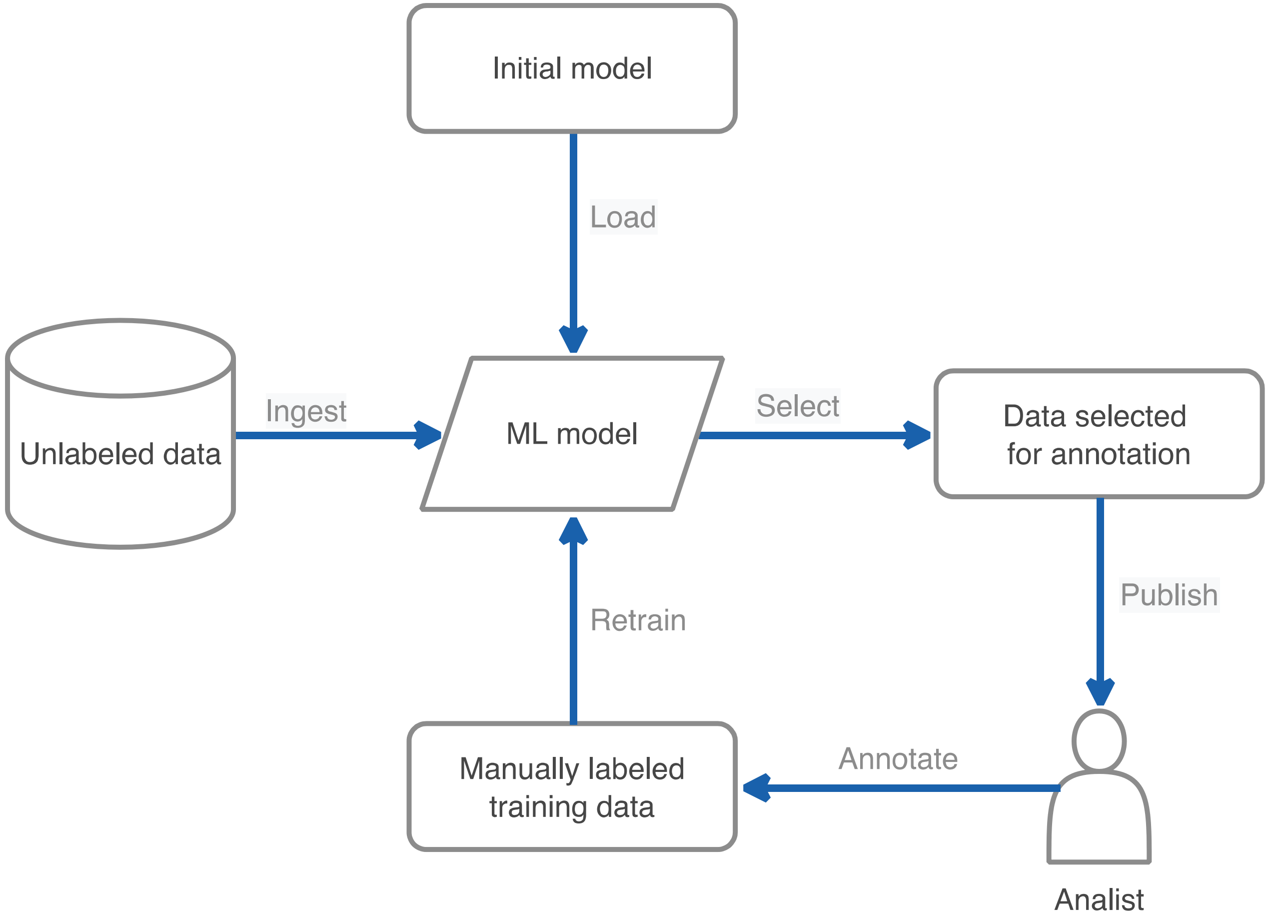


Part 4/4

Goals of active learning

- Minimize the labelling effort of human annotators
- Increase the accuracy of a machine learning model
- Reach the target accuracy of a machine learning model faster

Human-in-the-loop learning process

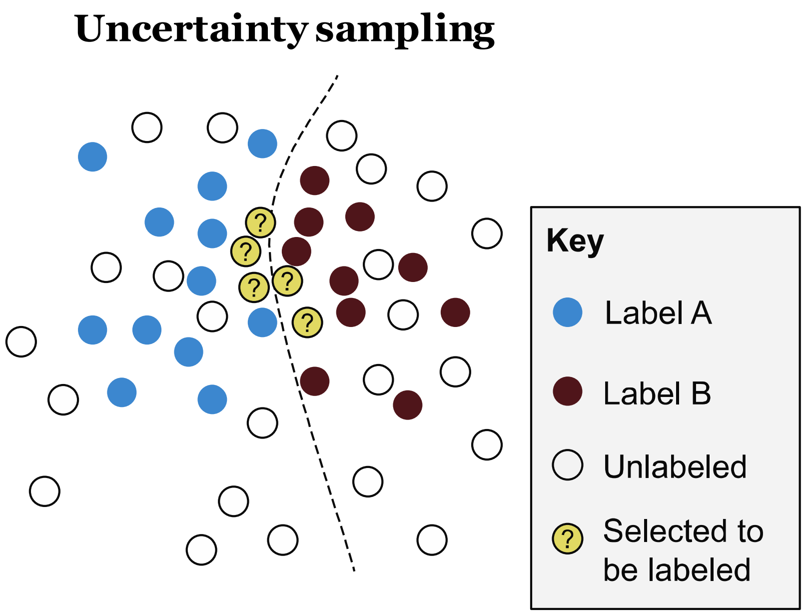


Active learning is no free lunch

- What is a relevant datapoint?
- What if the model assumptions are wrong?
- How many relevant datapoints should be labeled?
- Does model performance improve?

What is a relevant data point?

- Random sampling: each item has a fair chance of being selected (unbiased)
- Uncertainty sampling: select items close to decision boundary of a model
- Diversity sampling: select items underrepresented or unknown to a model
- Community disagreement sampling: select items that a community of models classify differently



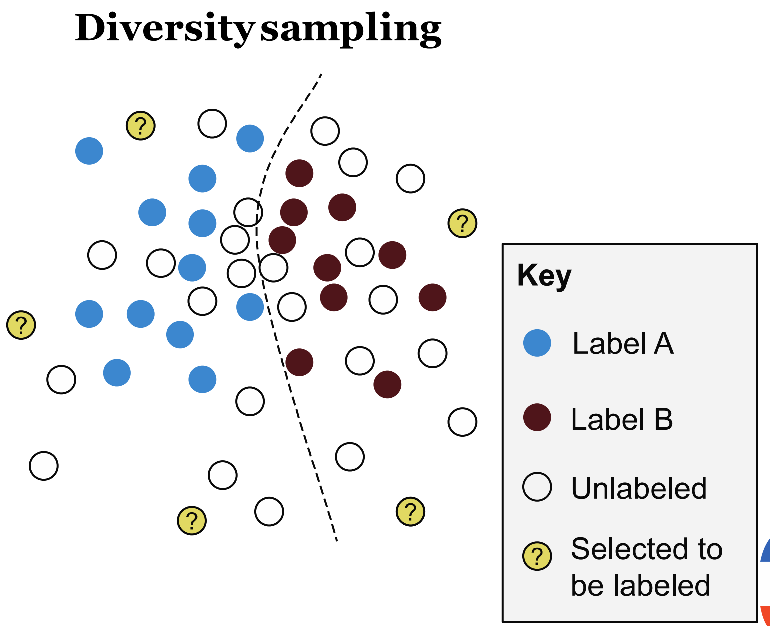


Figure source: *Human-in-the-Loop Machine Learning*, Robert (Munro) Monarch, Manning Publications.

DN LABS

Assignment 2: improve model using active learning

- Explore implemented sampling strategies
- Find best sampling strategy to improve model performance
  - A training iteration every week
  - Annotation budget: 50 data points per iteration
  - Measure improvement using average precision (AP)
- Implement your own sampling strategy (if time permits)

SIDN LABS

Instructions

1. Find a teammate

2. Browse to **github.com/SIDN/tma22_ml** and click on Assignment2 in the README

Results assignment 2

Go to www.menti.com and enter 3393 6819

Volg ons

SIDN.nl

@SIDN

SIDN

Q&A

www.sidnlabs.nl | stats.sidnlabs.nl

Thymen Wabeke
Research engineer
thymen.wabeke@sidn.nl

Thijs van den Hout
Research engineer
thijs.vandenhout@sidn.nl

Thanks to Unsplash.com and its photographers for beatifying these slides