Using machine learning to boost internet and DNS security

Thymen Wabeke

September 23, 2021

SIDN: operator of the .nl TLD

Registration of domain names

6.2M .nl-domains

Lookup IP address of a domain name

2.5B DNS queries/day

SIDN Labs: research team

- Goal: increase the trustworthiness (security, stability, resilience, and transparency) of our society's internet infrastructure, for .nl and the Netherlands in particular

- Strategies:
  - Applied 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, domain name & IoT security, trusted future internet infrastructures

Create ML applications that boost internet security

- Goal: apply ML to increase security of the Internet and DNS
  - Many large datasets are available
  - Manually extracting patterns is difficult

- 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)

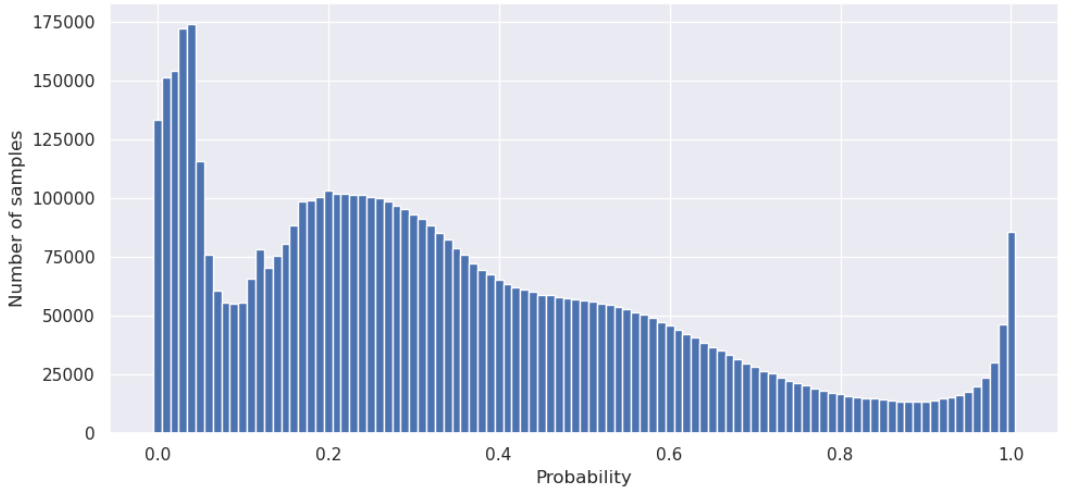
Applying ML in a responsible way

- Human-in-the-loop
- Simple and interpretable models
- Collaborate and publish
- Monitor performance

Radboud University

REALTIME REGISTER

Counterfighting Counterfeit: detecting and taking down fraudulent webshops at a ccTLD



UNIVERSITY OF TWENTE.

r a large array of goods, such as pharmaceu-

are among the
from the con-

ICS VISA MasterCard
INTERNATIONAL CARD SERVICES

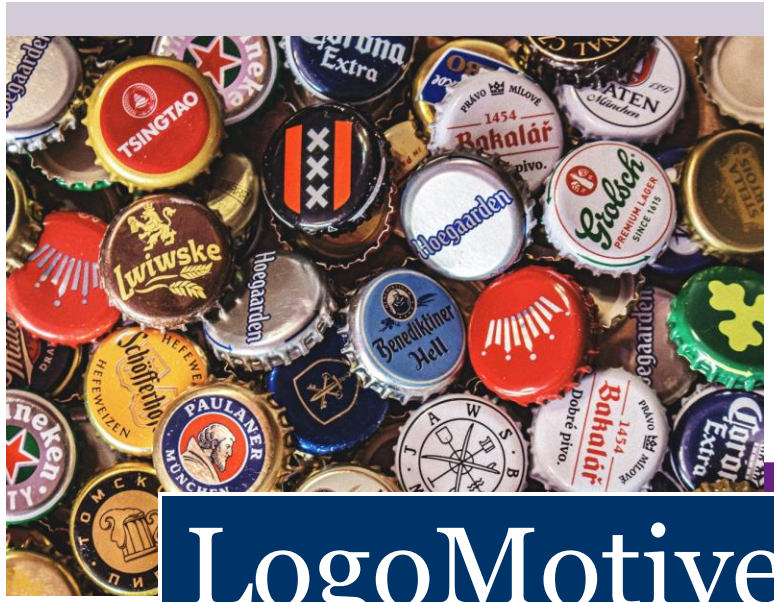
Thursday 10 June 2021
Article by: Thijs van den Hout, Thymen Wabeke, Cristian Hesselman

The original blog is in Dutch. This is the English translation.

SIDN LABS

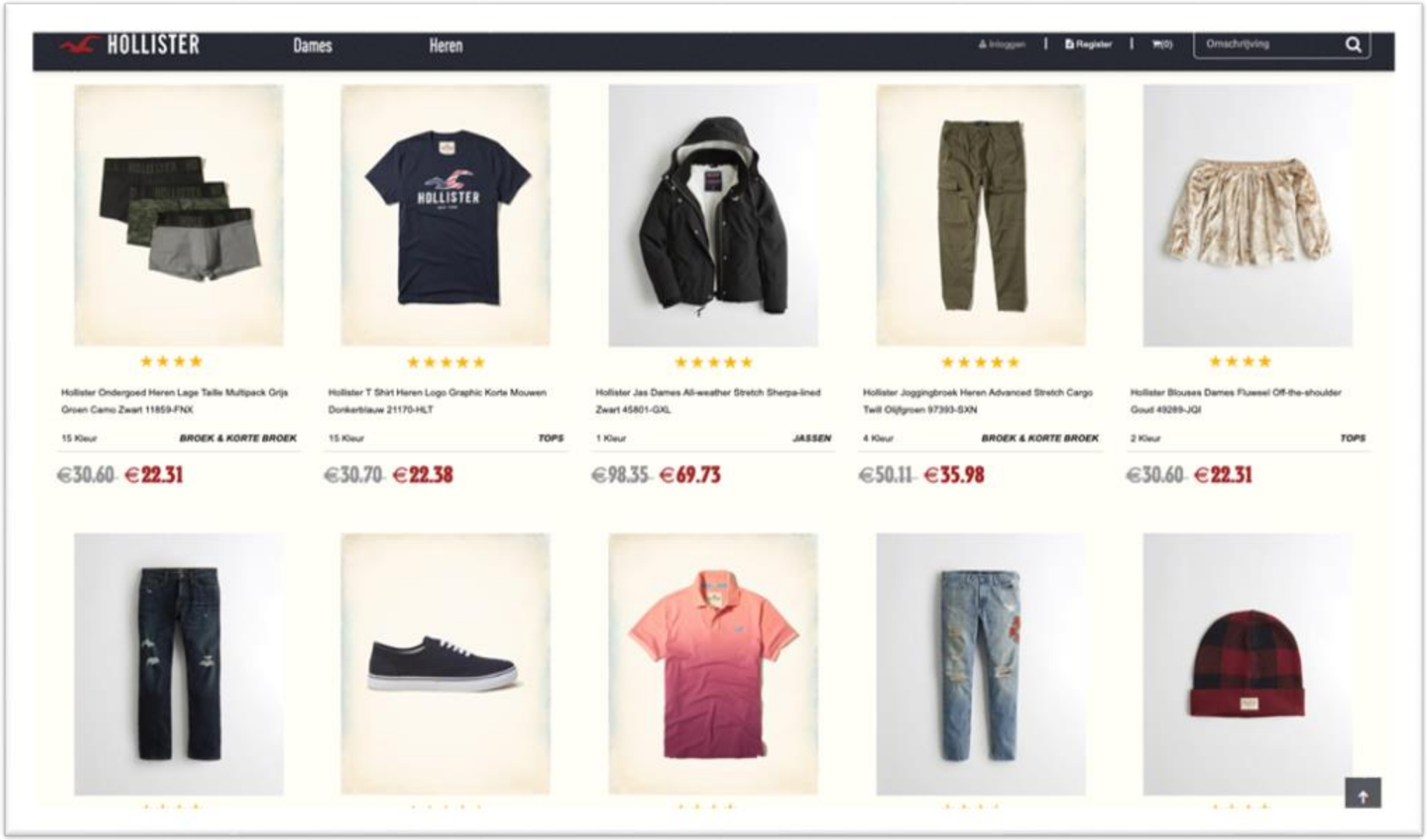
Remainder of presentation

FaDe



LogoMotive

nederlandwebshop.nl



SIDN's interest

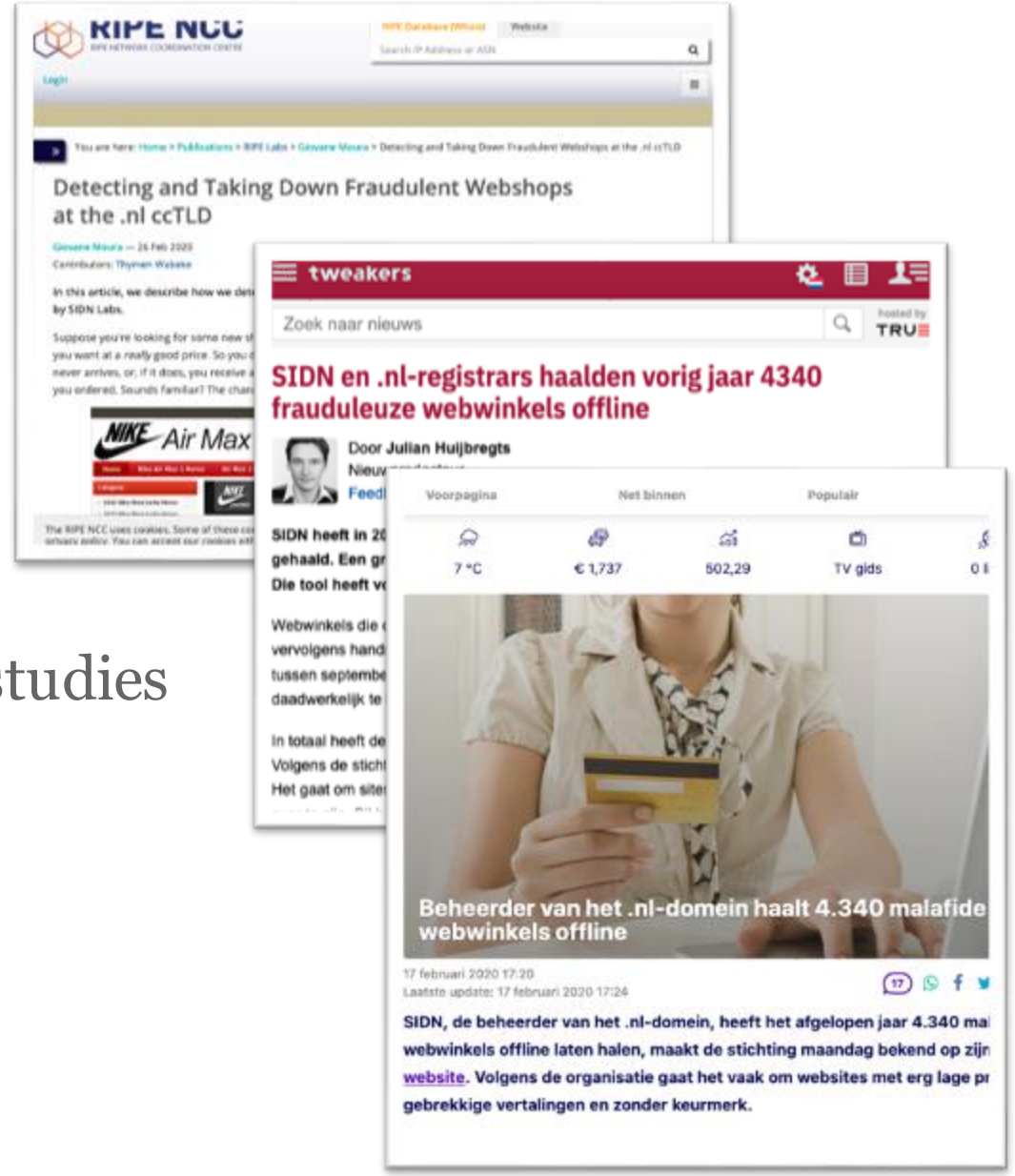
- Consumer losses
- Trust in Internet may decrease

Perfect vantage point:

- List of *all* .nl-domains, including:
  - DRS: Registration data
  - DMAP: DNS, TLS, SMTP and HTTP crawls
  - ENTRADA: traffic observed by authoritative name servers

Main results

- Detected thousands since 2016
- Protected users from being scammed



- PAM2020 paper: 2 detection systems, 2 case studies
  - BrandCounter (2018 Q1-2)
  - **FaDe (2019 Q1)**

Published at PAM2020: https://doi.org/10.1007/978-3-030-44081-7_10

Fake Detector (FaDe)

- Collaboration with ICS Cards, a credit card issuer
- ICS Cards provided 231 shops involved in scams
- Classification model based on supervised machine learning

Photo by Sigmund on Unsplash

- Support Vector Machine
- Optimized using grid search

Dataset

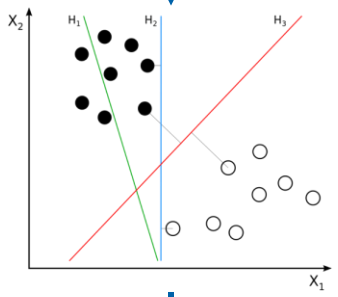
- 231 counterfeit
- 229 legitimate

Features

- 6 registration
- 3 infrastructure

Training samples

Train model



Testing samples

Apply model

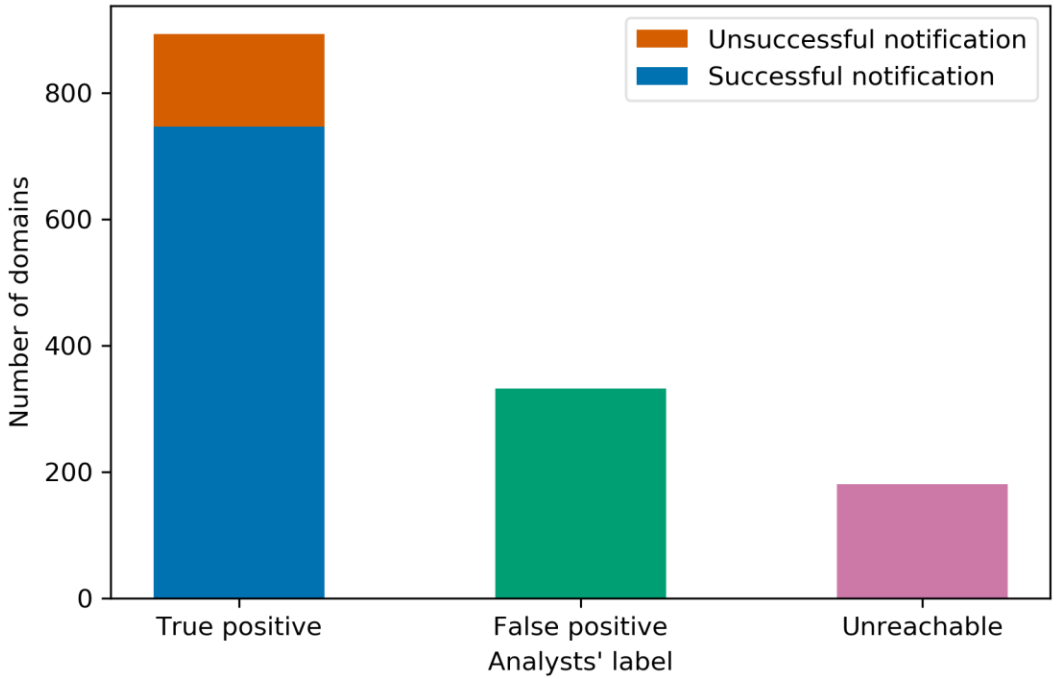
Samples	Precision	Recall
Train (cross-validation)	0.98	0.97
Test	1.0	1.0

SIDN LABS

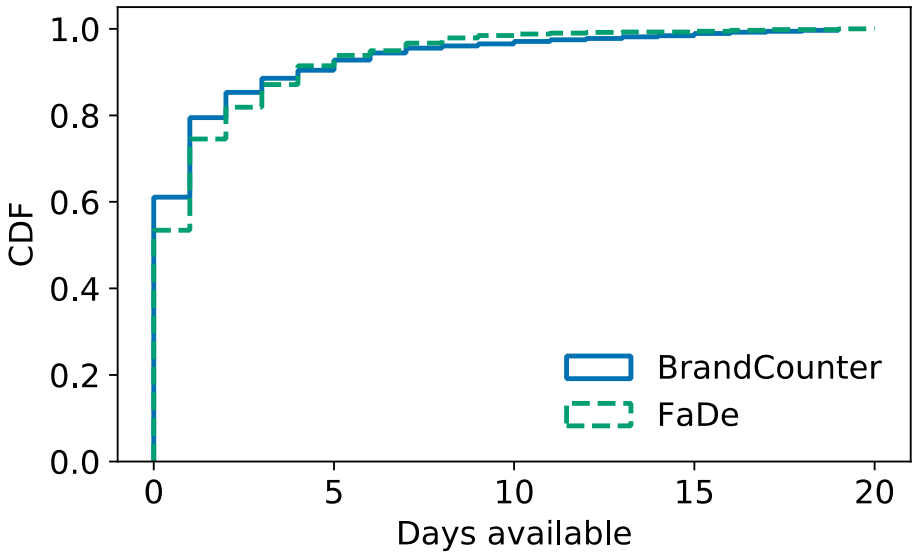
FaDe notification

- Applied model to 30k .nl-domains
  - 1407 classified as suspicous
  - 894 true positives (73%)

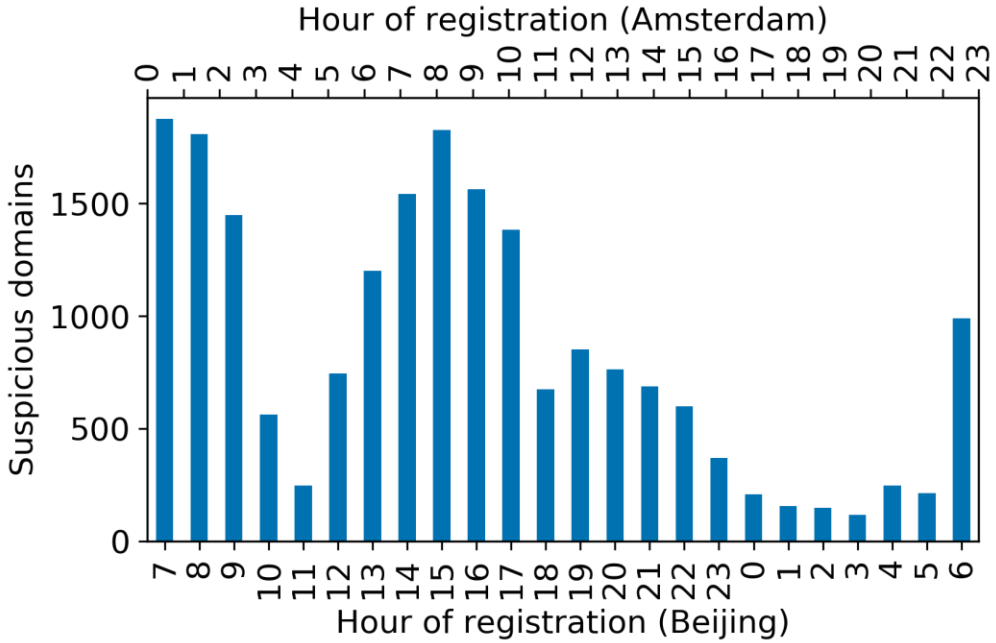
- Sent 894 notifications to registrars
  - 747 taken down (84%)



How do counterfeiters operate?



Days between domain expiration and re-registration



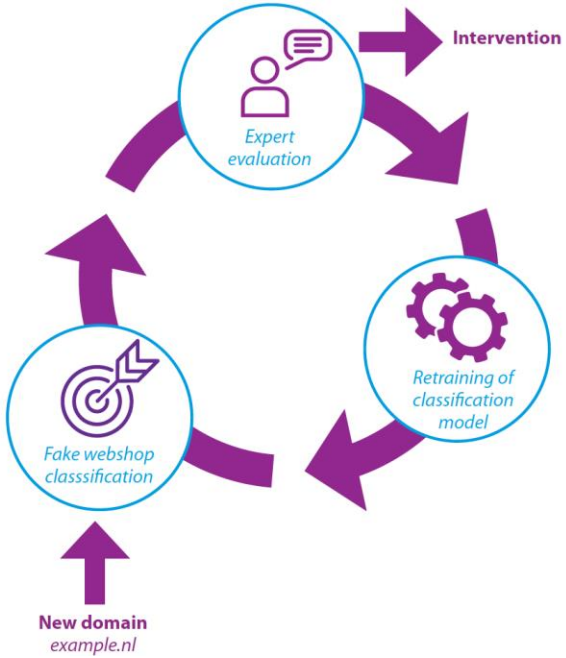
Time of domain registration

Lessons learned

- Registrars and ICS collaboration was key

- Detectors are simple yet effective
  - Registries have perfect vantage point
  - Suggests little pressure

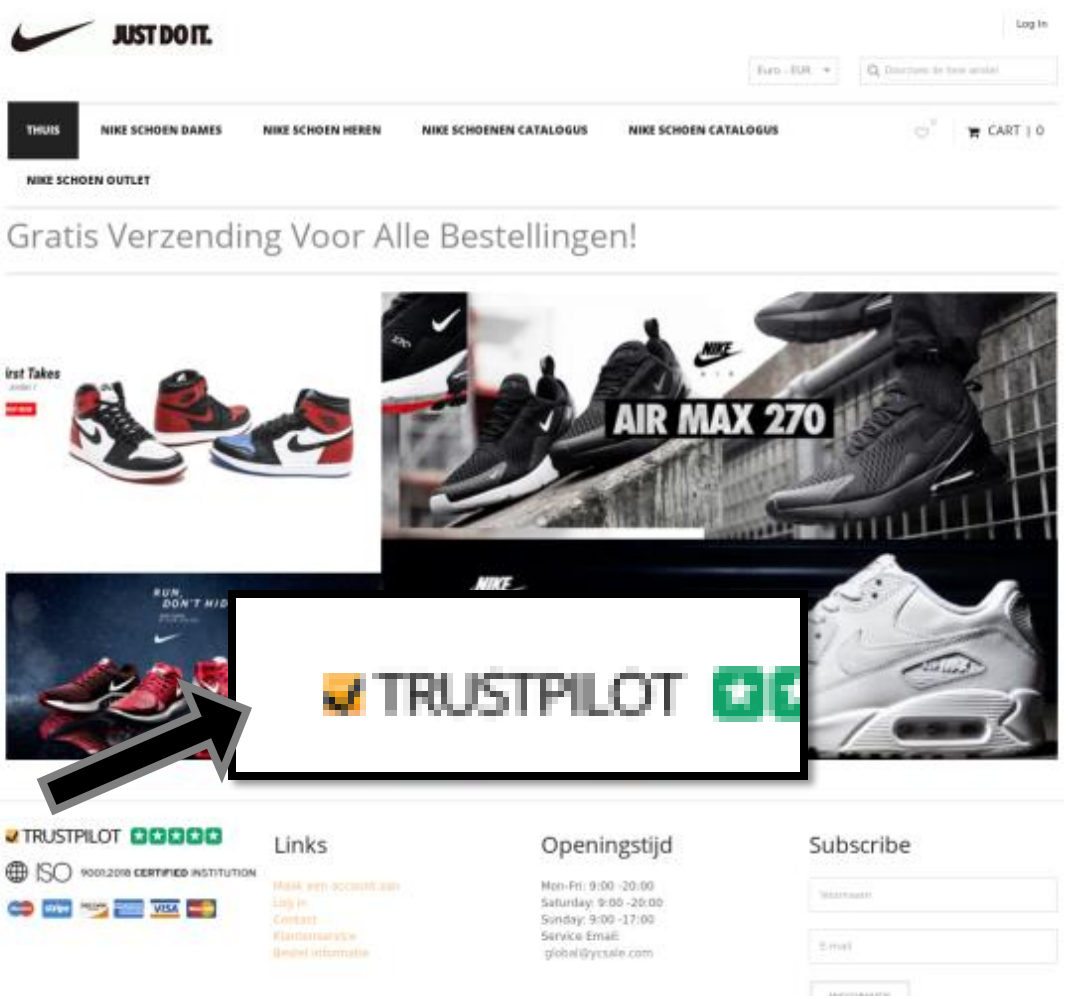
- It's an ever-going whack-a-mole game
  - Monitor features and evaluate model regularly
  - Fewer takedowns = fewer scams?



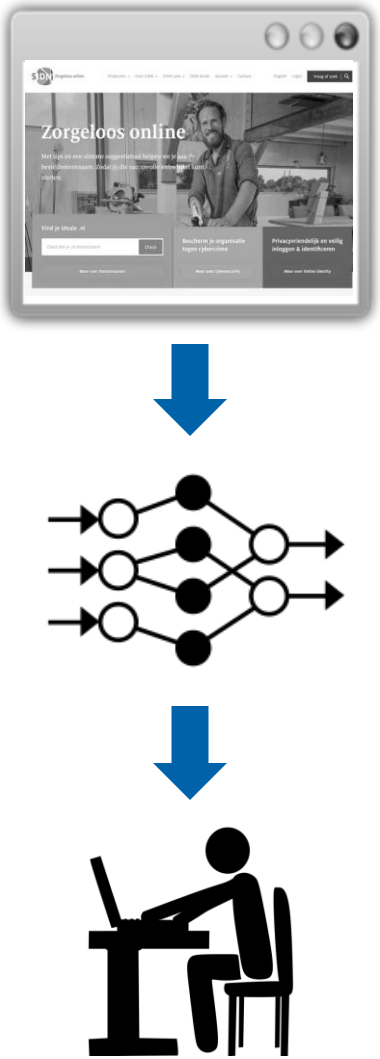
Year	Taken down
2018	~12,000
2019	4,340
2020	481

Number of counterfeit webshops taken down

Malicious websites use well-known organizations' logos



Help analyst find suspect websites using logo detection



1. Crawl and take screenshots

- Should be efficient: skip duplicates
- Using Selenium

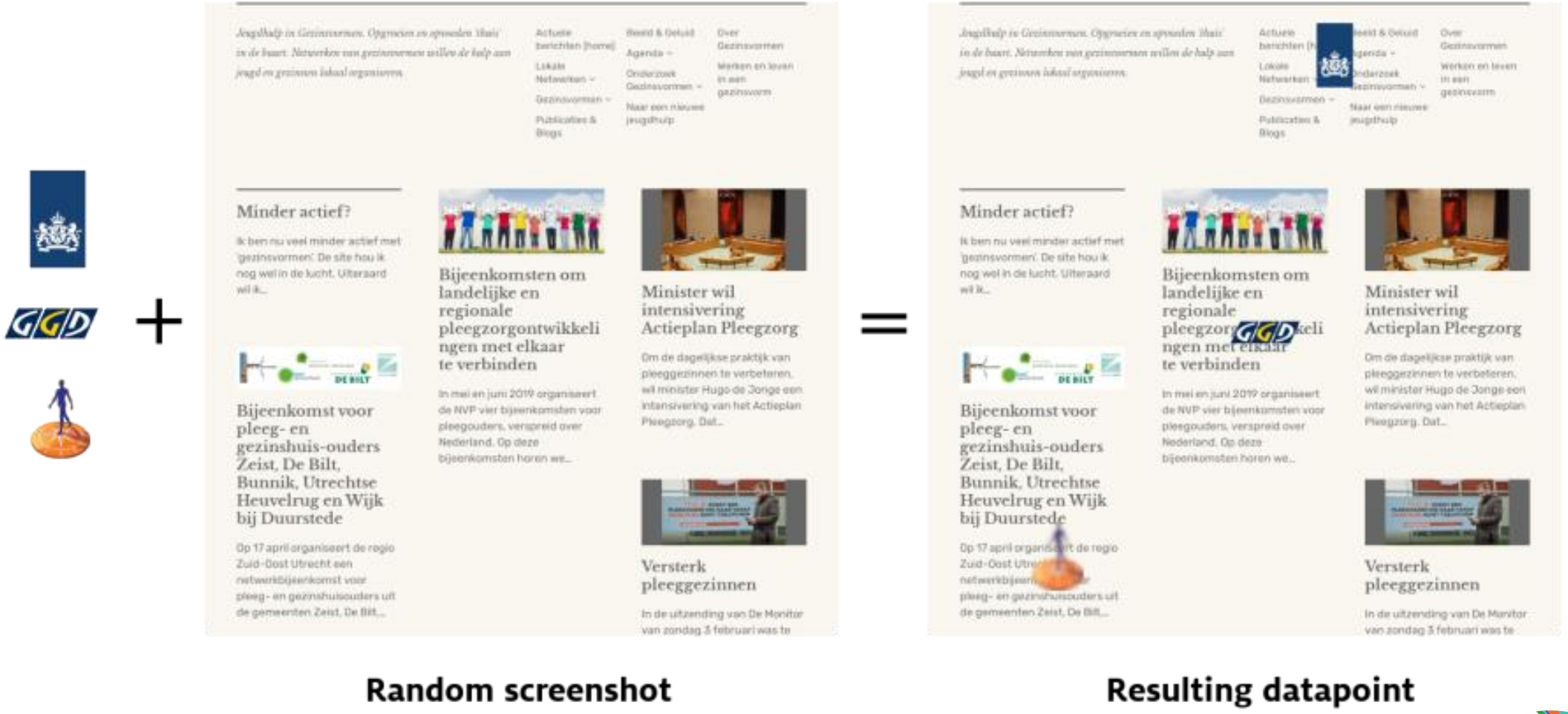
2. Object detection algorithm

- Should be flexible: no manual labeling
- Based on YOLOv5

3. Annotation dashboard

- Should be easy-to-use and generic

Automatic training data generation



Random screenshot **Resulting datapoint**

Filter

Label All

Status All

Filter

Bulk update

Label Select

Status Select

Update

Logo's found

Show 10 entries Select All Search:

Domain name	Screenshot date	Registrar	Registrant	Registered on	Label	Status	
sidn.nl	2021-08-12 11:11	Stichting Interne...	Stichting Interne...	1999-11-18	-	Open	Annotate
domainregistry.nl	2021-08-12 11:11	Stichting Interne...	Stichting Interne...	2001-03-27	-	Open	Annotate
dnsops.nl	2021-08-12 11:11	Stichting Interne...	Stichting Interne...	2007-09-26	-	Open	Annotate
dnssec.nl	2021-08-12 01:11	Stichting Interne...	Stichting Interne...	2001-03-26	-	Open	Annotate
6miljoen.nl	2021-08-11 22:08	Stichting Interne...	Stichting Interne...	2020-06-12	-	Open	Annotate
sidnlabs.nl	2021-08-11 10:18	Stichting Interne...	Stichting Interne...	2010-05-10	-	Open	Annotate
abuse204.nl	2021-08-11 01:39	Stichting Interne...	Stichting Interne...	2014-08-11	Correct use	Afgehandeld	Annotate

Logo found on sidn.nl

Screenshot date	12-08-2021 11:11
Page also found on	dnsops.nl, schijtbakkes.nl, domainregistry.nl, vdstbv.nl
Registrant	Stichting Internet Domeinregistratie Nederland
Registrar	Stichting Internet Domeinregistratie Nederland 2
Registration date	18-11-1999 00:00
Screenshots	

Clear label

Previous

Comment

Comment...

Label

- Correct use
- Incorrect use
- Geen logo

Status

- Open
- In behandeling
- Afgehandeld

Save and update all related domains

Save and next

Save and exit

Can we use logo detection to boost internet security?

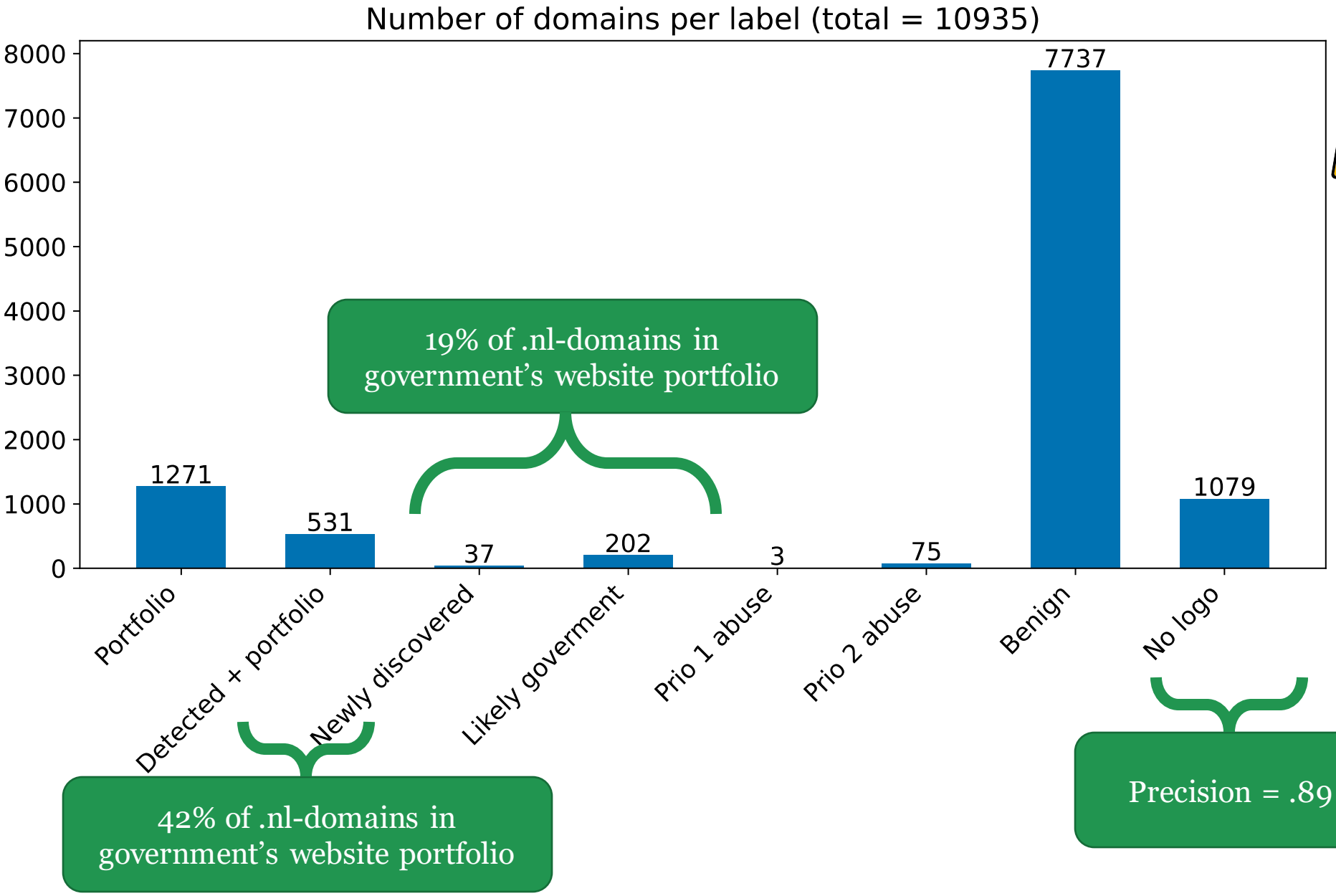
Apply to 6.2M .nl-domains and evaluate with  thuiswinkel waarborg

- Detect malicious webshop by matching results with member list
- Out of scope for today ☺

Apply to 6.2M .nl-domains and evaluate with  Rijksoverheid

- Detect abusive domains
- Discover unknown government domains

SIDN LABS


Number of domains per label (total = 10935)
8000
7000
6000
5000
4000
3000
2000
1000
0
1271
531
37
202
3
75
7737
1079
Portfolio
Detected + portfolio
Newly discovered
Likely goverment
Prio 1 abuse
Prio 2 abuse
Benign
No logo
19% of .nl-domains in government's website portfolio
42% of .nl-domains in government's website portfolio
Precision = .89


WORK
IN PROGRESS

Example detected abusive webpages

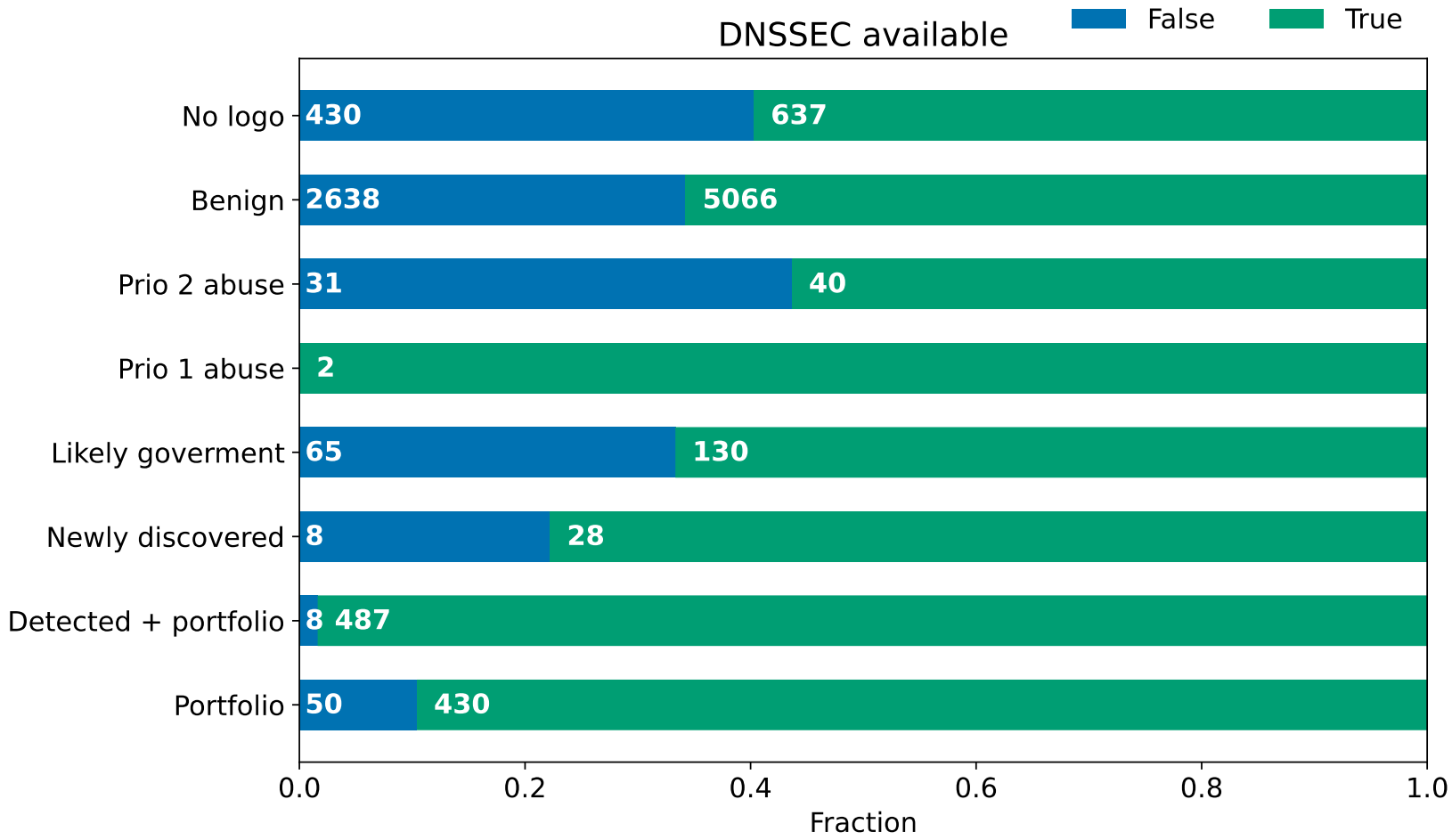
Prio 1

Prio 2

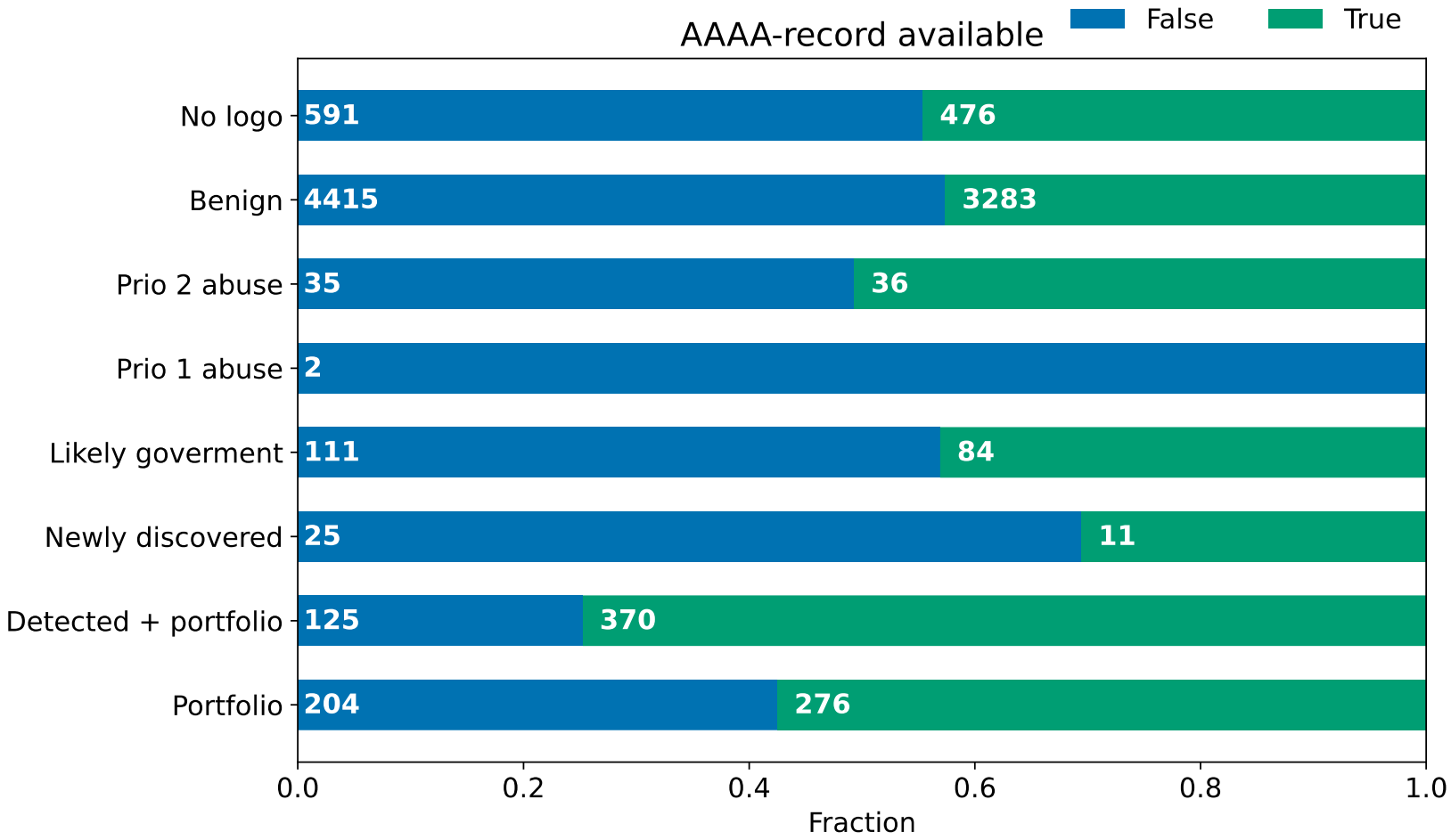
Importance of accurate domain portfolio

- Prevent leakage if domain got canceled by accident
- Full control over domain name and its DNS records
- Adopt and monitor security standards and technologies

DNSSEC adoption per label



IPv6 adoption per label



Lessons learned and future work

- Visual aspects like logo's help to detect abuse
- Logo's also help to keep domain portfolio accurate
- Large gray area of unwanted, but not abusive content

Next steps:

- Monitor new registrations
- Write academic paper and publish code

Volg ons

SIDN.nl

@SIDN

SIDN

Q&A

www.sidnlabs.nl | stats.sidnlabs.nl

Thymen Wabeke
Research engineer
thymen.wabeke@sidn.nl