Counterfighting Counterfeit

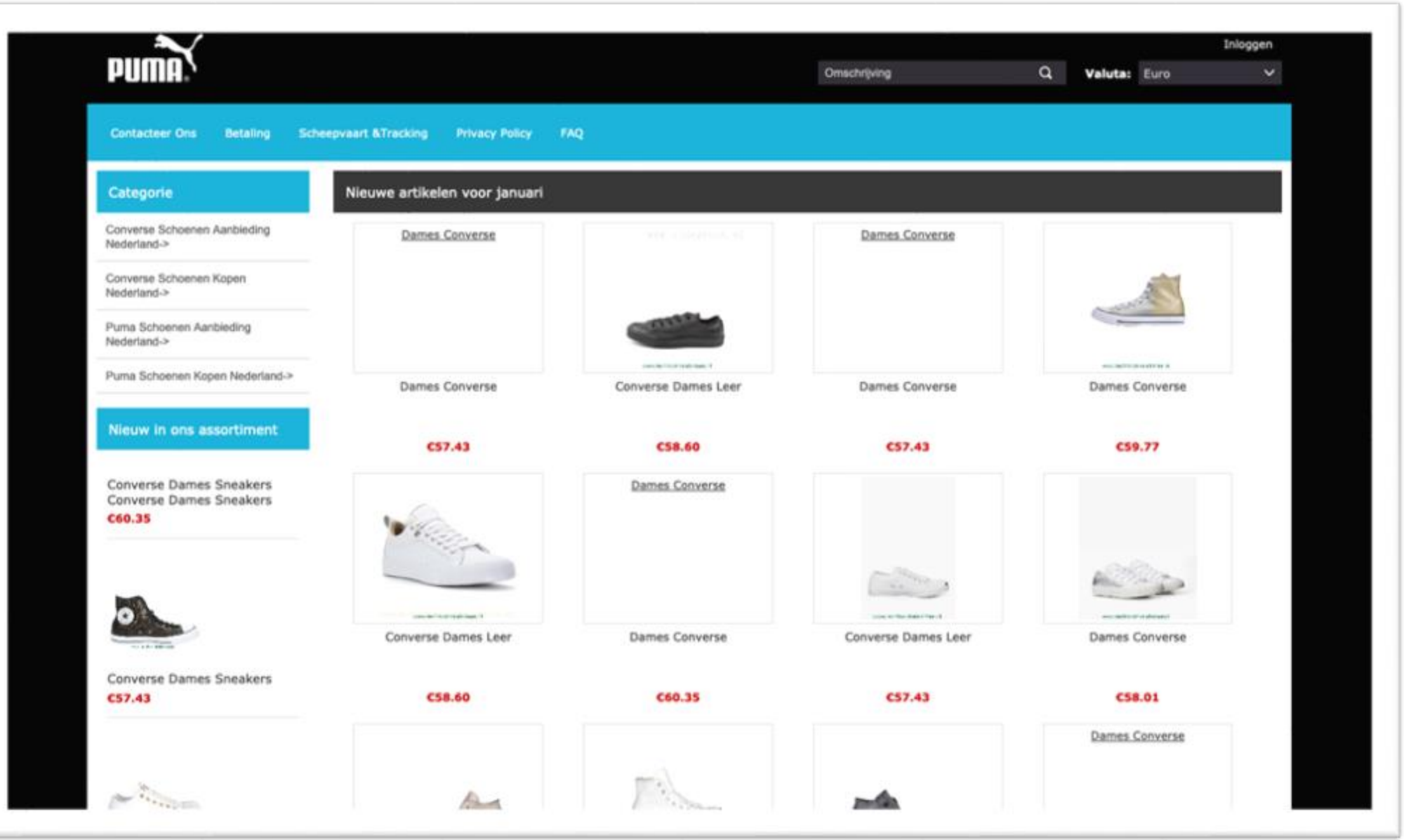
Detecting and taking down fraudulent webshops

Thymen Wabeke | RON++

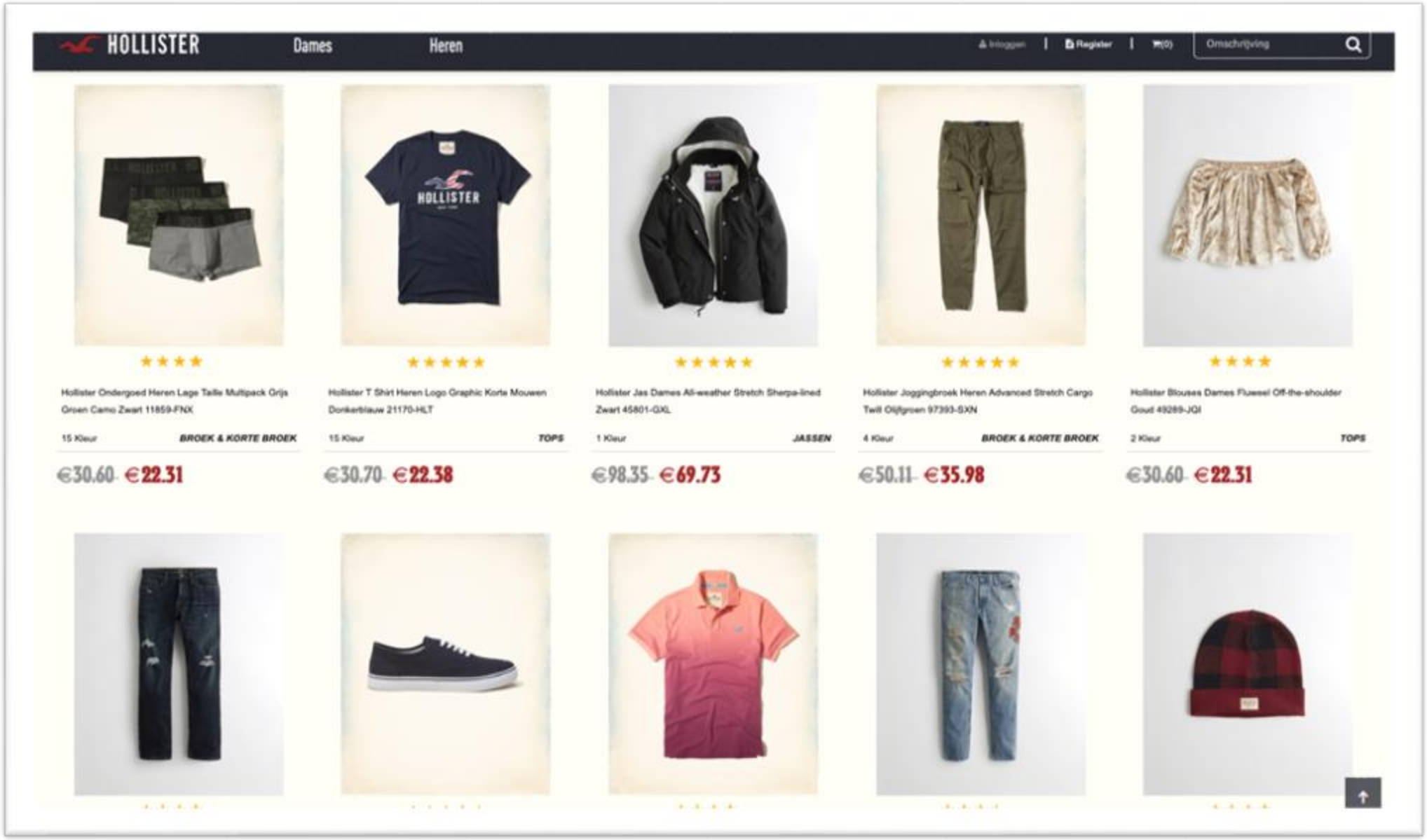
SIDN LABS


NIKE AIR
AIR FORCE

rechtswinkelalkmaar.nl



nederlandwebshop.nl



Who? Why?

- Consumer demand for counterfeit goods
  - $1.2 billion seized at US border in 2017
  - €670 million seized at EU border in 2016
- Low investments, easy money
- Less attention by law enforcement

https://www.cbp.gov/document/stats/ fy-2017-ipr-seizure-statistics | https://ec.europa.eu/taxation_customs/ node/976_en

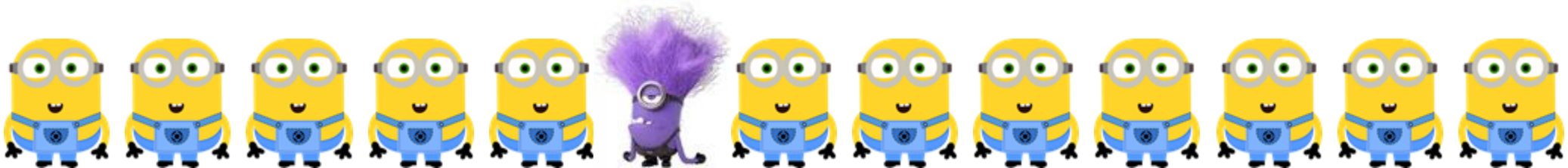
FAKE

Rabobank

ICS VISA MasterCard
INTERNATIONAL CARD SERVICES

How to detect fake webshops at scale?

(.nl has over 5,9 million domain names)



2016

- BrandCounter
- Study with registrar (~3.7k domains removed)

2018

- Student project machine-learning based detection

2019

- Study with International Card Services (747 domains removed)
- **Development of FaDe (2088 domains removed)**

2020

- **Publication at PAM2020 ☺**
- **Future research…**

FaDe: a robust fake detector

Adaptive

Pro-active

Accurate

Photos: *Marc A on Unsplash* | *TPapi on Flickr* | *Sonja Langford on Unsplash*

Adaptive by extracting rules

Knowledge-driven programming:

Data-driven programming:

Source: *Deep Learning with Python* (François Chollet)

Machine-learning ingredients

Answers (i.e. ground truth)**:**

- Examples found during earlier studies (469, Sept ‘19)

Data (i.e. features)**:**

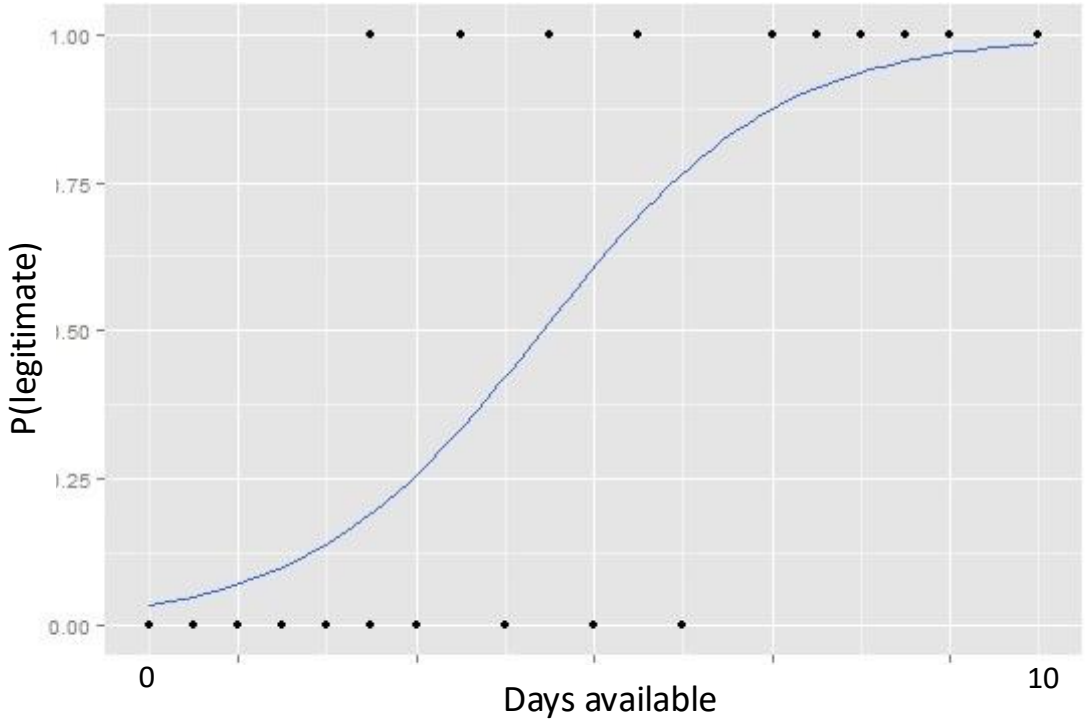
- WHOIS (registrar, registrant e-mail, re-registration)
- Content (html title, description, page type)
- Infrastructure (AS, SMTP configured, TLS issuer)

Algorithm:

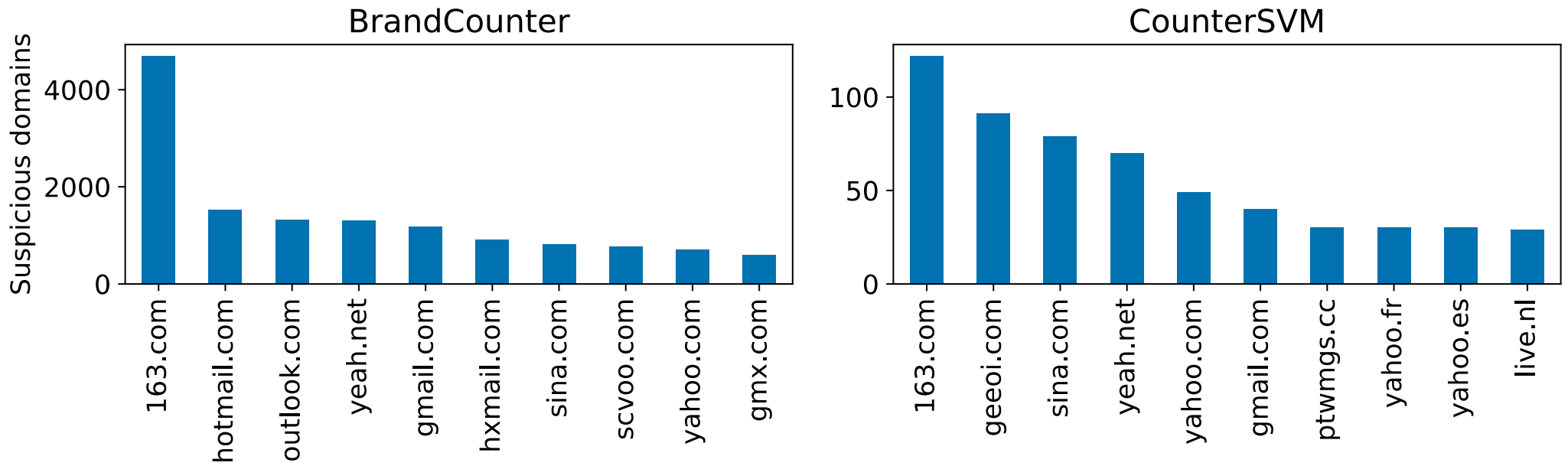
- Logistic regression

Why logistic regression?

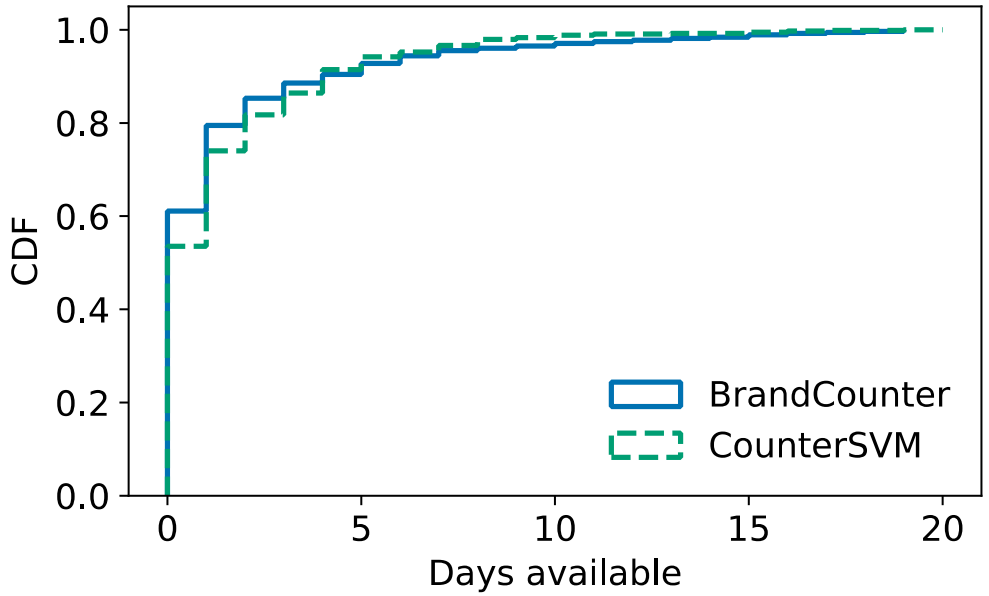
- Linear: interpretable by default
- Calibrated: inferences have confidence measure
- Simple: less chance of overfitting, computes fast



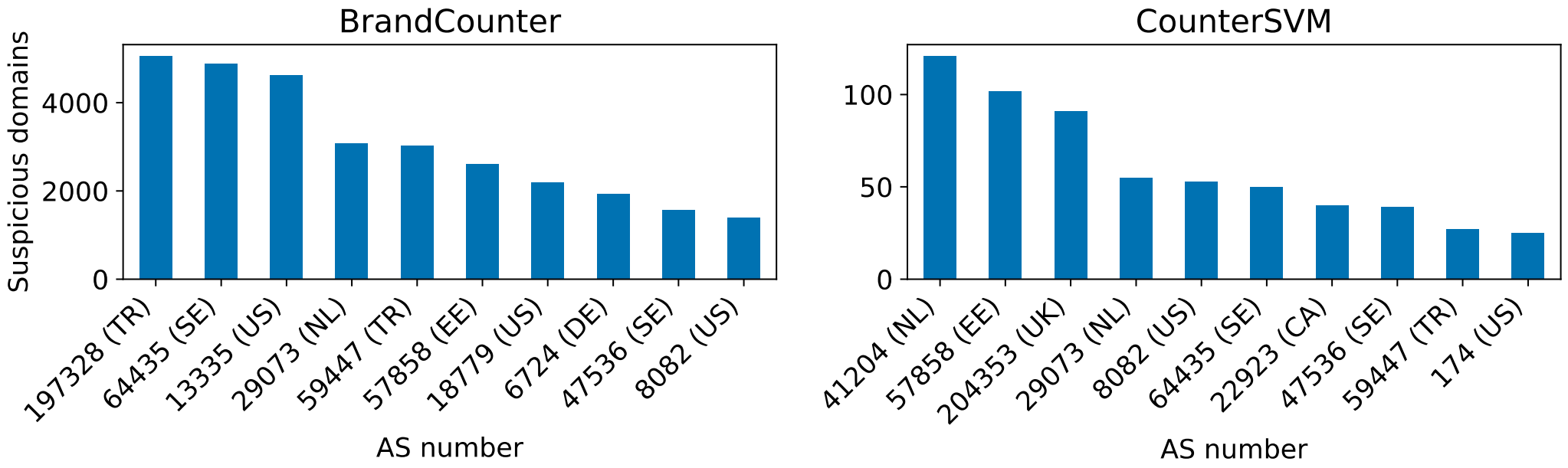
Registrant e-mail domain



Days before registration



Hosting AS


BrandCounter
Suspicious domains
4000
2000
0
197328 (TR)
64435 (SE)
13335 (US)
29073 (NL)
59447 (TR)
57858 (EE)
18779 (US)
6724 (DE)
47536 (SE)
8082 (US)
AS number
CounterSVM
100
50
0
41204 (NL)
57858 (EE)
204353 (UK)
29073 (NL)
8082 (US)
64435 (SE)
22923 (CA)
47536 (SE)
59447 (TR)
174 (US)
AS number

Detect before they harm

Fake Webshops

This page lists the domain names that are found by Padawans.

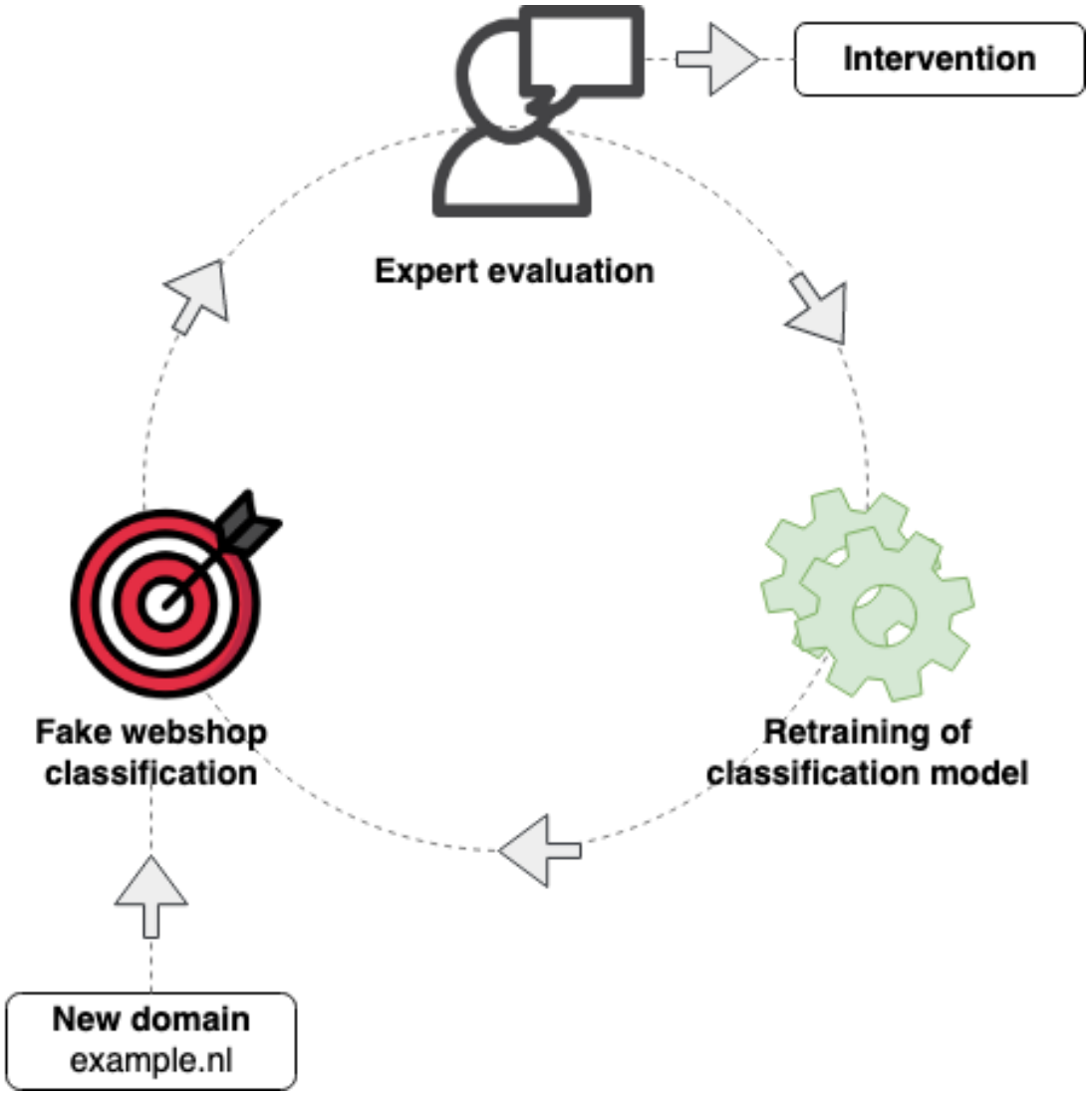
Show 100 entries

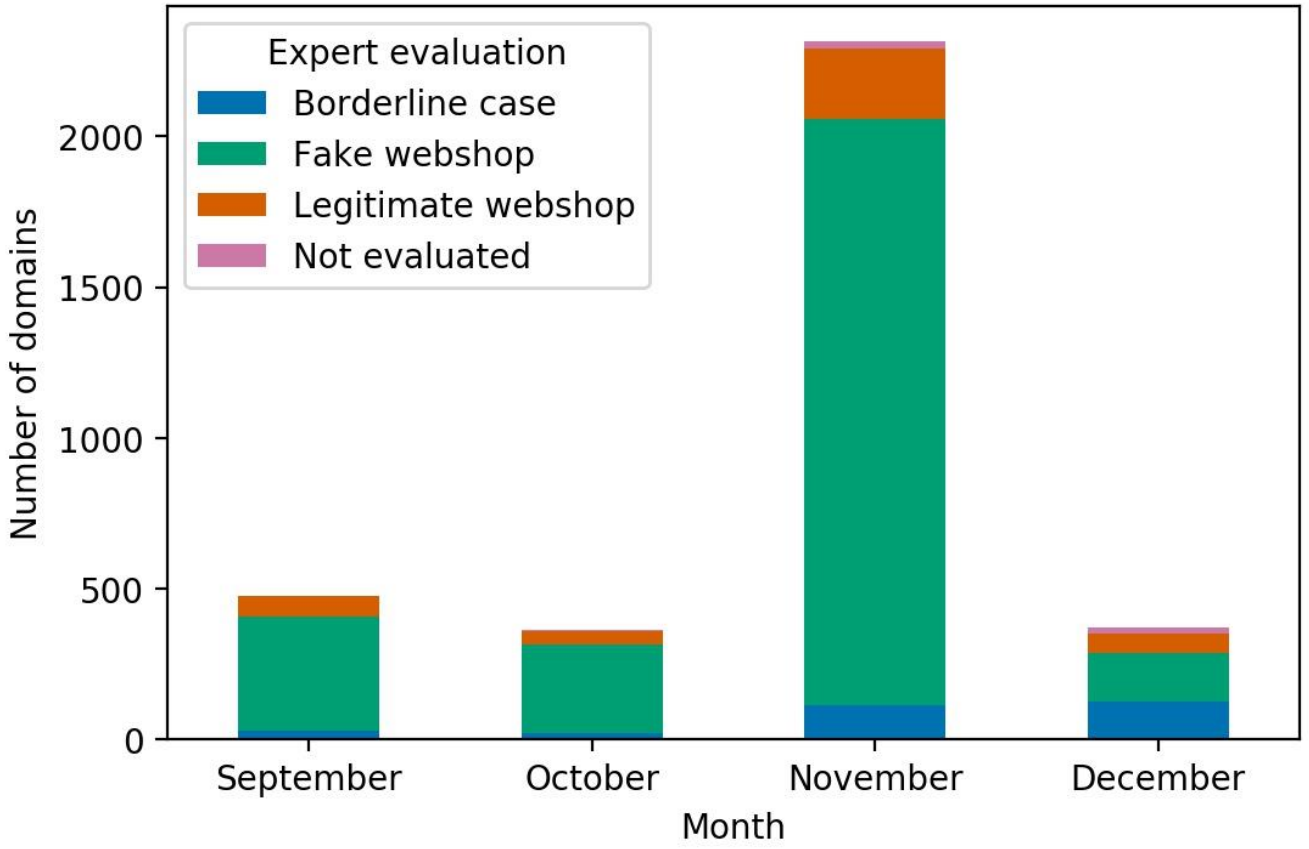
Search:

Copy CSV

Domain Name	Added	Registrar	Registrant	Mail	Feedback	Submit Feedback	Link	Phone	Whois
echo[redacted]	2020-01-03 08:04:51.432482	[redacted]	[redacted]	[redacted]com		Fake|Valid|Unknown	click	+49.0[redacted]	click
npot[redacted]	2019-12-26 08:23:01.263207			[redacted]gmail.com		Fake|Valid|Unknown	click	+84.3[redacted]	click
tank[redacted]	2019-12-23 08:08:08.768590			[redacted]acyprotect.org		Fake|Valid|Unknown	click	+45.3[redacted]	click
only[redacted]	2019-12-23 08:07:45.236334			[redacted]@gmail.com		Fake|Valid|Unknown	click	+86.1[redacted]	click
nike[redacted]	2019-12-23 08:07:43.051879			[redacted]nxmail.com		Fake|Valid|Unknown	click	+86.3[redacted]	click
mtkf[redacted]	2019-12-23 08:07:42.009627			[redacted]1@gmail.com		Fake|Valid|Unknown	click	+1.67[redacted]	click
mrw[redacted]	2019-12-23 08:07:41.722979			[redacted]acyprotect.org		Fake|Valid|Unknown	click	+45.3[redacted]	click
miko[redacted]	2019-12-23 08:07:39.971509			[redacted]@163.com		Fake|Valid|Unknown	click	+93.1[redacted]	click
bega[redacted]	2019-12-23 08:07:06.592282			[redacted]acyprotect.org		Fake|Valid|Unknown	click	+45.3[redacted]	click
drm[redacted]	2019-12-21			[redacted]163.com		Fake|Valid|Unknown	click	+49.0[redacted]	click

Learn from feedback





	Notified	Registrars	Removed by registrar	Art 16 by SIDN
Dec ‘19	2091	3	2088	3

November peak: scam season

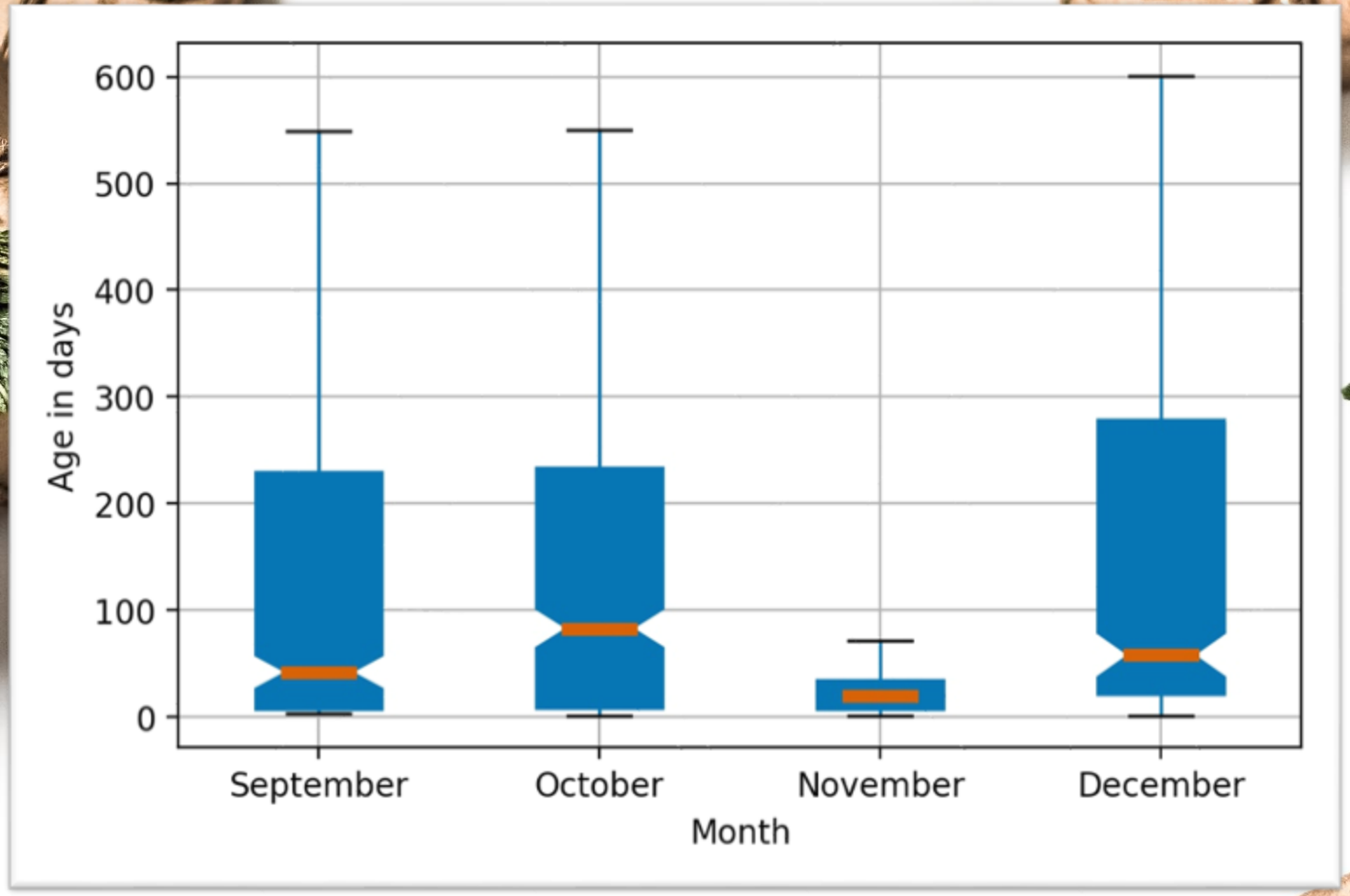
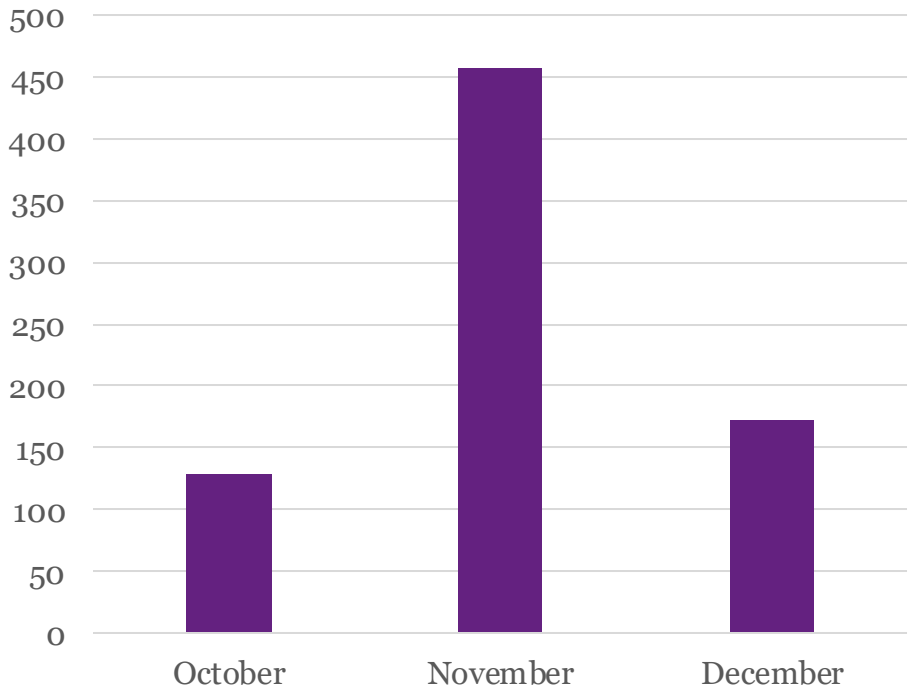


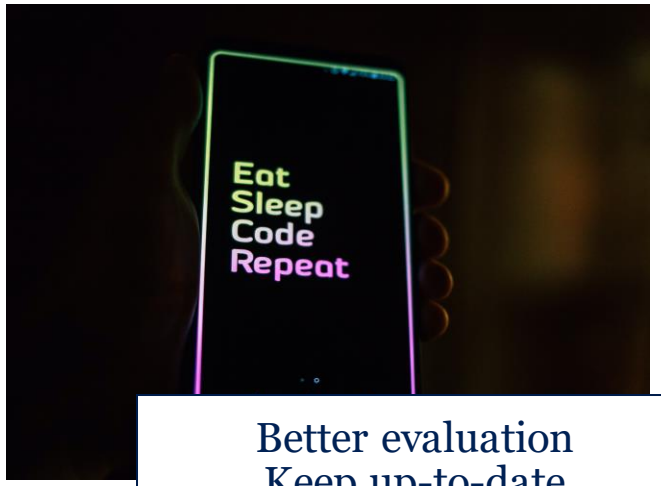
Photo: Caley Dimmock on Unsplash

November peak: better model

- 458 found in November, overlooked in October
- 173 found in December, overlooked in November

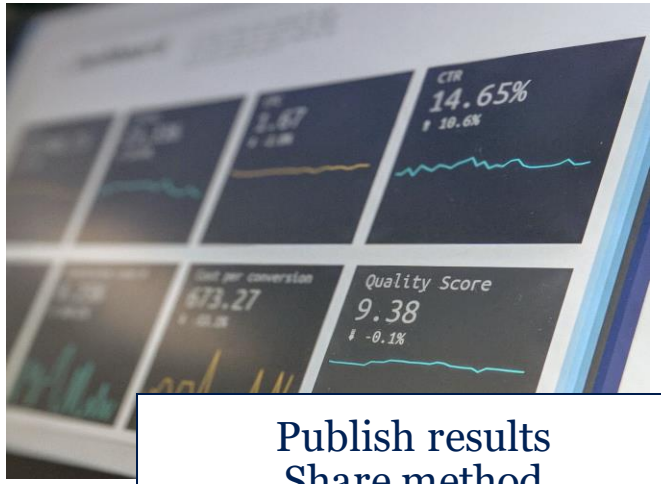


Next steps

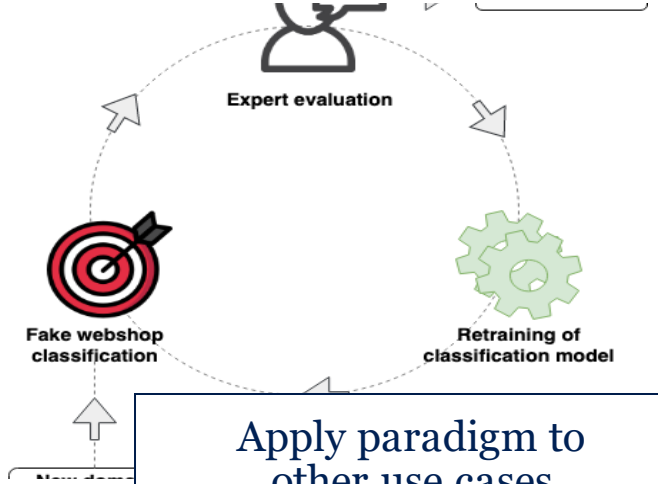


Better evaluation
Keep up-to-date

Identify campaigns
Follow the money
Effective interventions



Publish results
Share method



Apply paradigm to other use cases

Photos: Roman Synkevych, Stephen Dawson and Michael Longmire on Unsplash