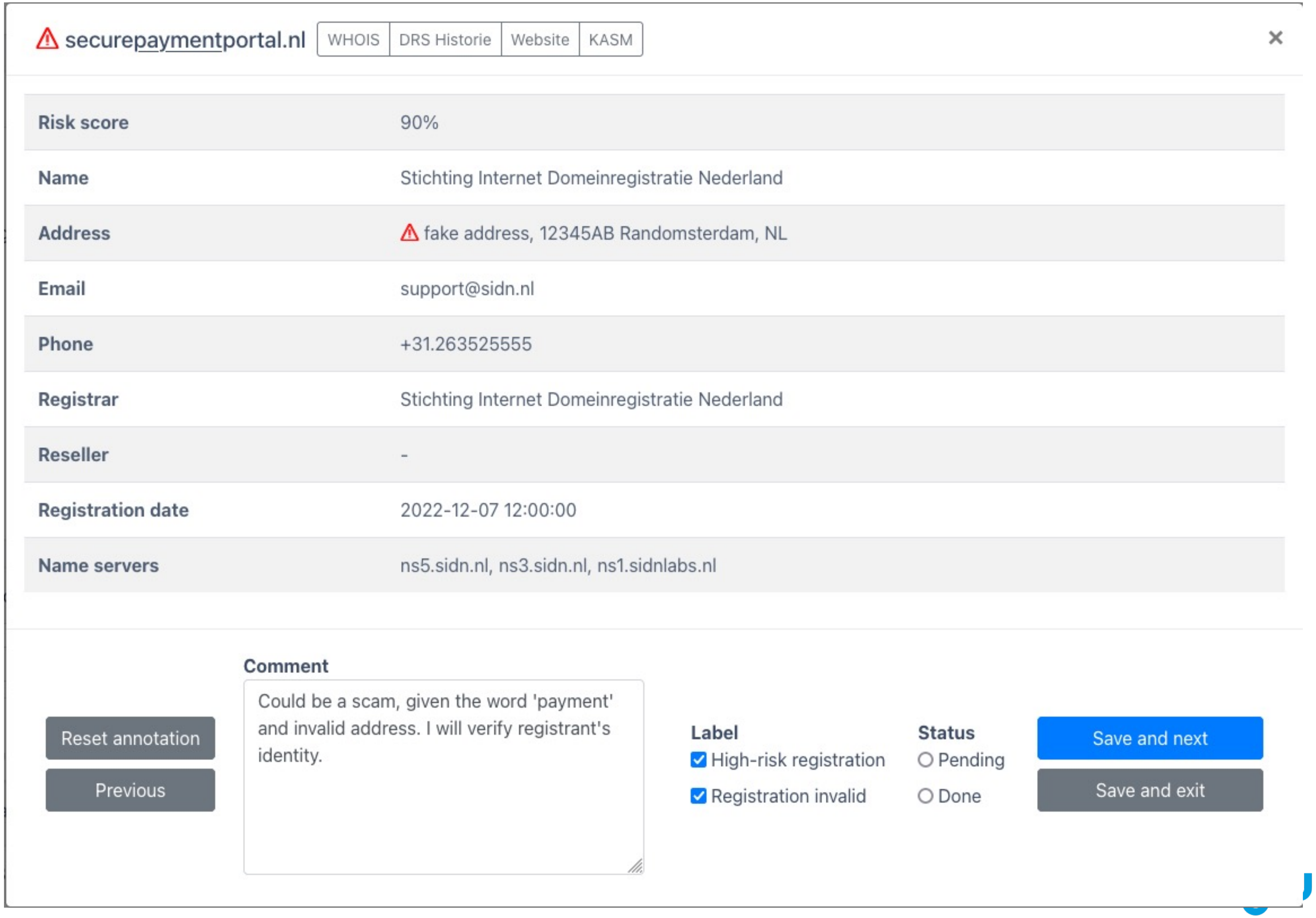
Identification of potentially malicious registrations

Maarten Bosteels (.be), Thijs van den Hout & Thymen Wabeke (.nl) | 22nd CENTR R&D

Agenda

- Status update SIDN & DNS Belgium
- Plans for joint project
- Discussion

Status update SIDN



Our year in a nutshell

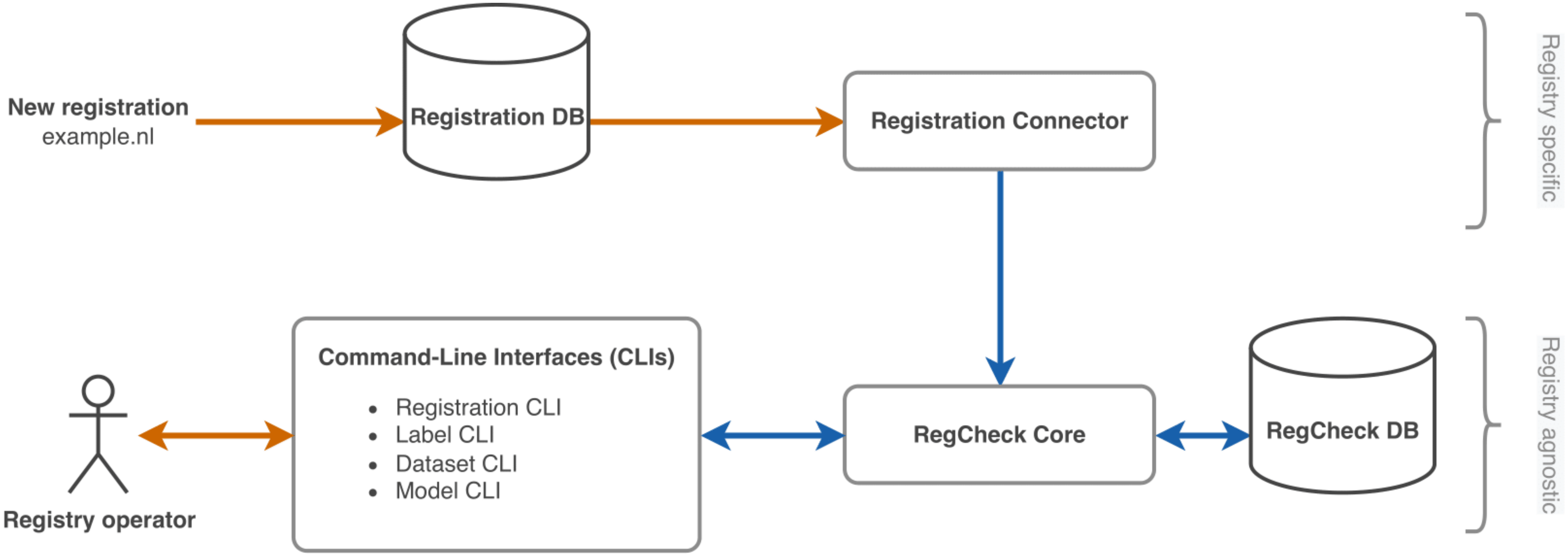
20nd R&D meeting

- Results of feasibility study
- "PhD code"
- Registry specific

22nd R&D meeting

- Operational results
- Mature system
- Registry agnostic
- Interpretable risk scores

RegCheck design



Classifiers that calculate risk scores

- Calculation by looking at risk factors individually
  - Risk factor: feature that increases a registration's risk (11 currently)
  - Advantage: you can interpret risk scores
  - Disadvantage: you cannot model nonlinear relations (does not seem like a big deal)

- Rule-based and machine learning classifier
  - Feature constructors and classifiers follow scikit-learn's interfaces
  - Advantage: you can use scikit-learn utilities, such as `Pipeline` and `GridSearchCV`

Offline and online results

	Machine learning	Rule based
Recall	48%	9%
PPV (precision)	22%	0.55%

Table 1: RegCheck's results on historical data (August to November 2022).

	Machine learning
Registrations	43k
High-risk classifications	181 (0.4%)
True positives	38 (21%)

Table 2: RegCheck's results on new registrations (17 November to 8 December 2022).

Plans for 2023

- Continue discussion on response and registrant verification process
- Embed RegCheck with NIS2 measures
- Help other registries by sharing our code
- Joint project with DNS Belgium…

Photo by Jess Bailey on Unsplash

Status update DNS Belgium

- Rule-based system in production since November 2020
- Configurable to a certain degree (keywords, threshold, …)
- If registration is selected
  - ➢ Delegation delayed
  - ➢ Registrant needs to prove his identity
- Around 15% of new registrations
- Plan to ramp up to 100% set on hold (workload)
- Machine Learning to the rescue

Labels can be combined in several ways

~~Ground Truth~~

Weak Labels

IS BAD WHOIS	count	pct
True	27,836	2.58%

IS MALICIOUS	count	pct
True	17,706	1.64%

Same WHOIS data was used in a previous malicious registration

Domain name contains critical keyword (e.g., bank name)

No detected incidents 1 year after registration

Registrant verification procedure was started and is still pending

Training Labels

True

False

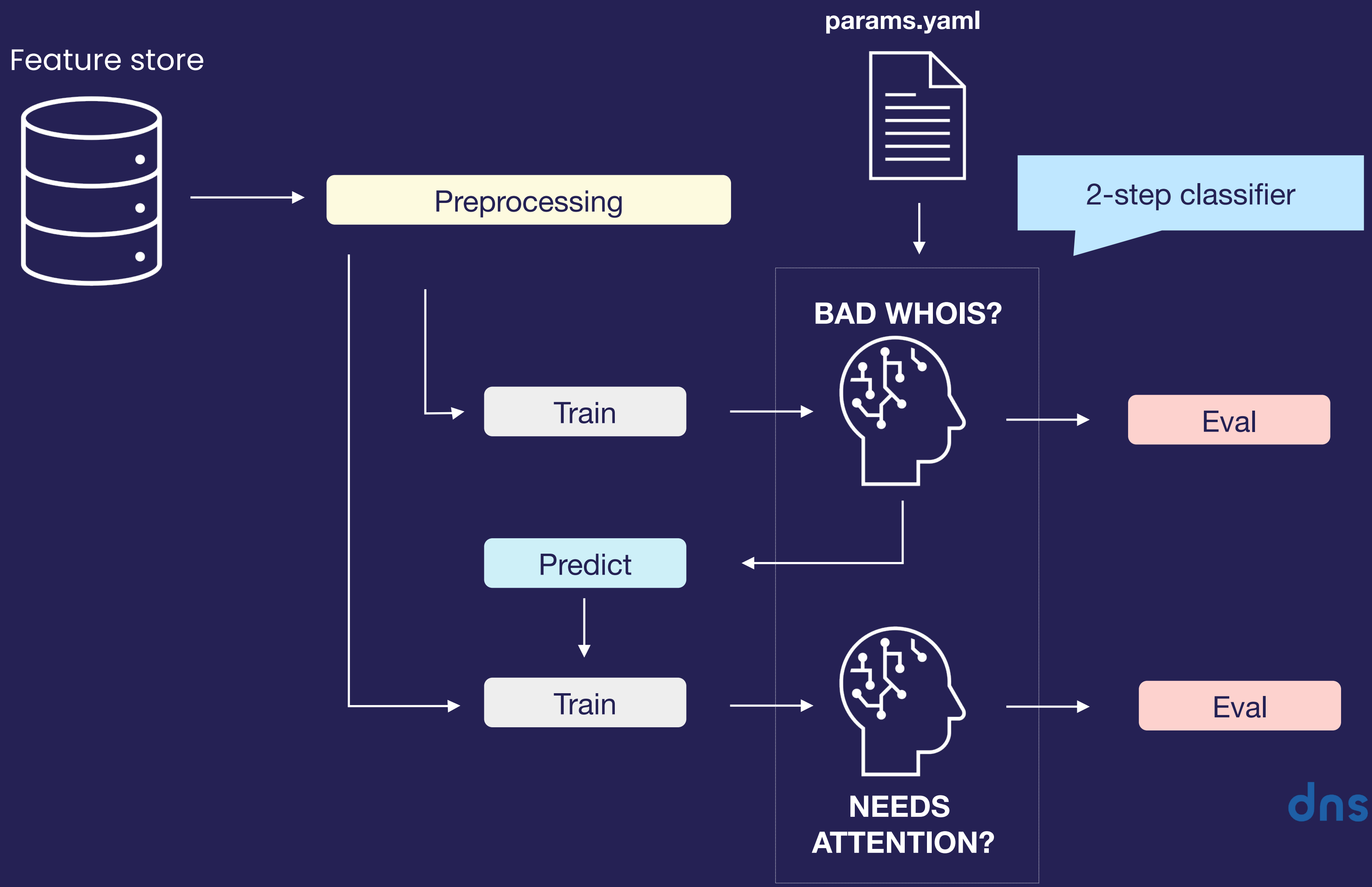
is_bad_whois

True

needs_attention

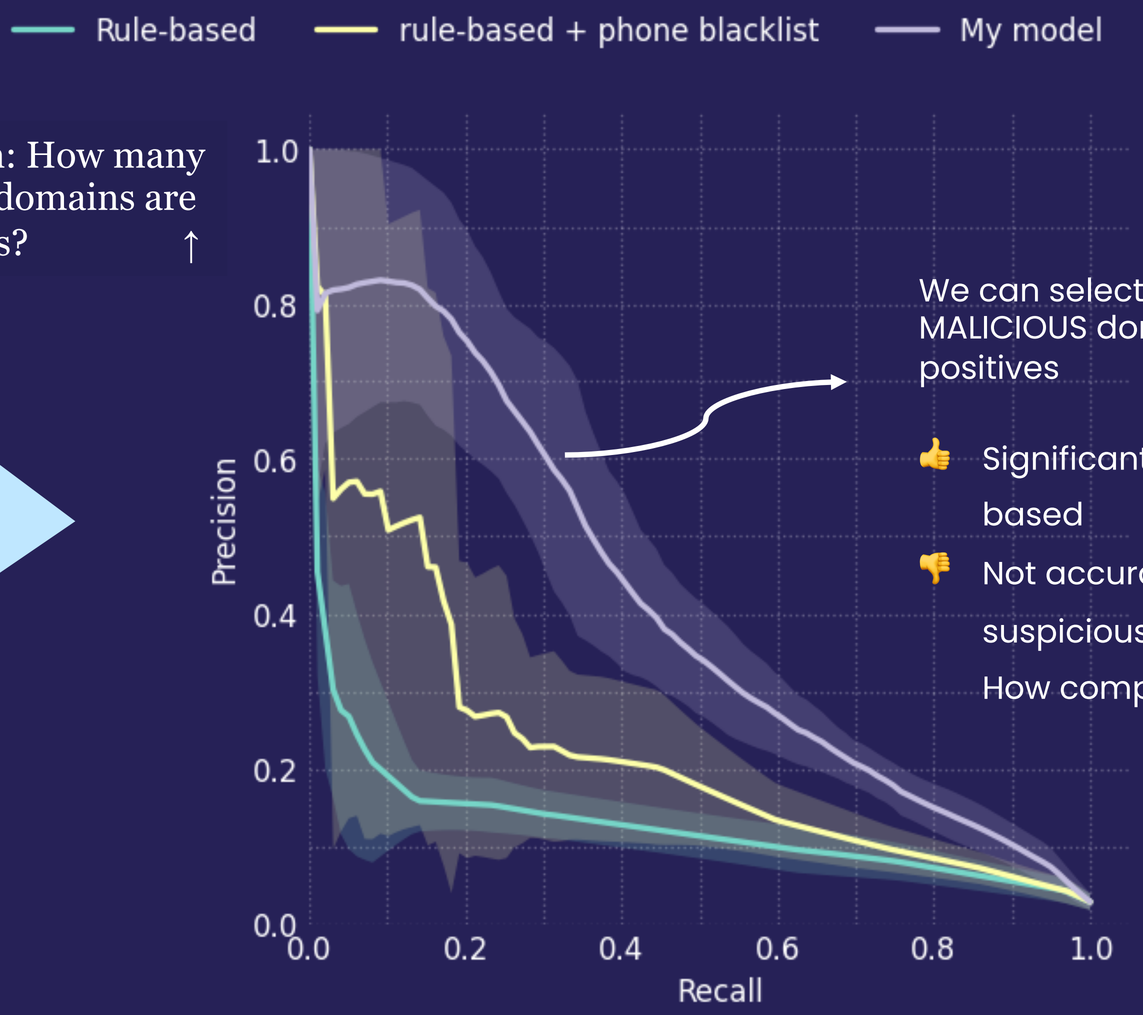
False

Machine Learning Pipeline



Needs Attention Classifier

Precision: How many selected domains are malicious? ↑



Trained on weak labels

Evaluated on the Nostradamus ground truth:

IS_MALICIOUS OR IS_BADWHOIS

We can select 30% of the BAD WHOIS / MALICIOUS domains, at the cost of 40% false positives

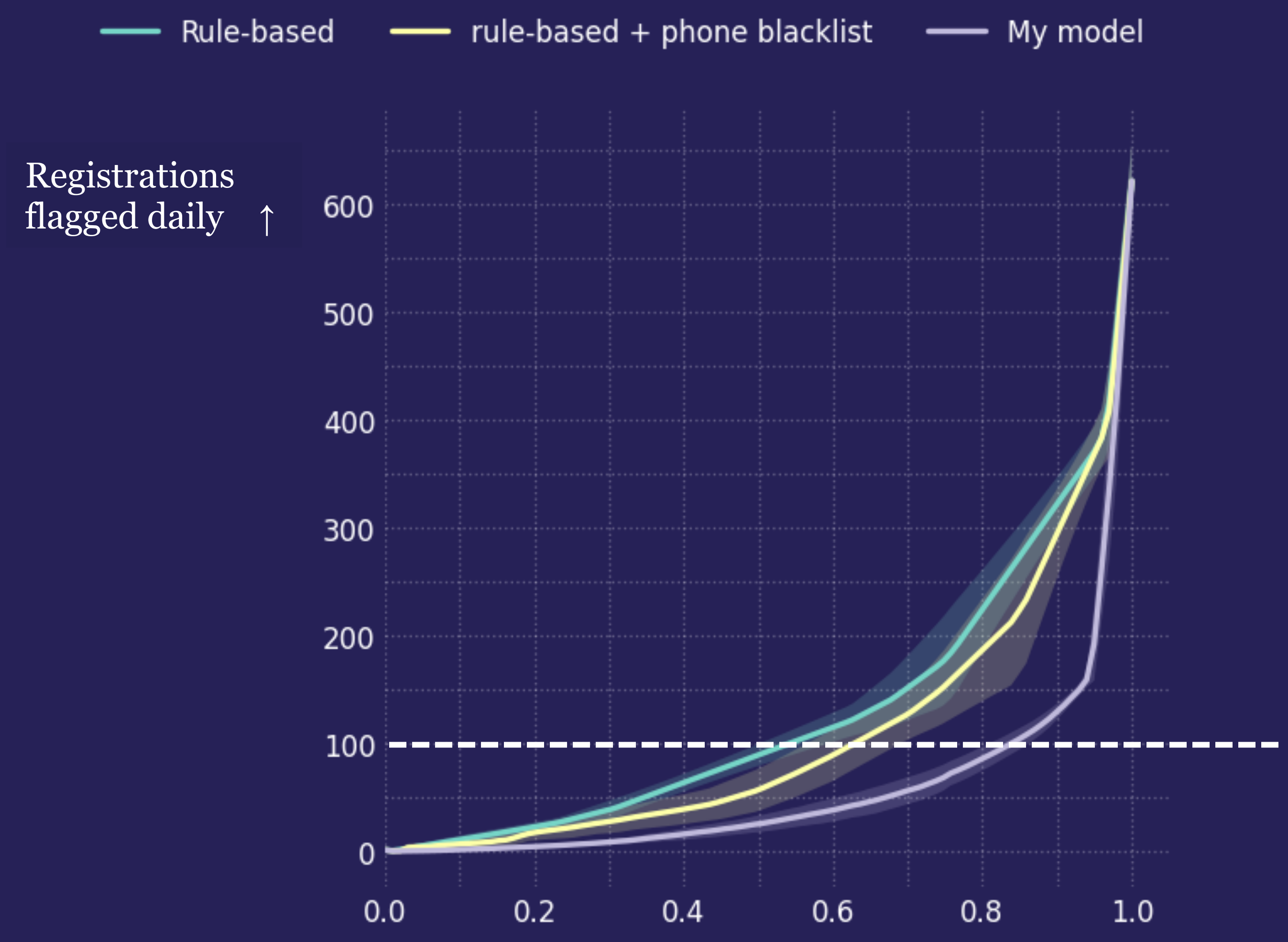
- 👍 Significantly more accurate than rule-based
- 👎 Not accurate enough to blindly reject all suspicious registrations

How complete is the ground truth?

Recall: How many malicious domains are selected? →

Data: Jan 2021 – Mar 2022

Needs Attention Classifier



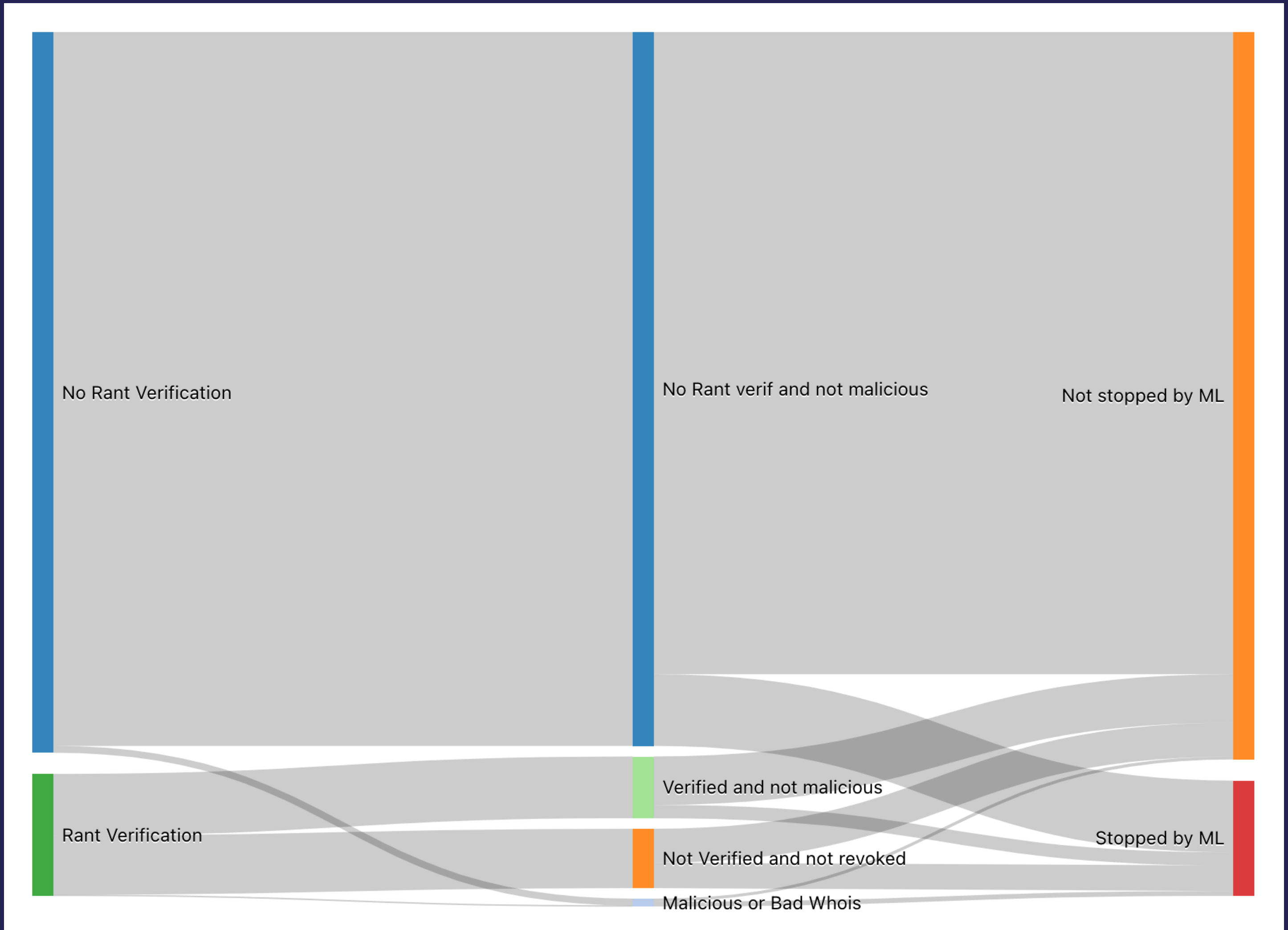
With 100 registrations flagged daily

- The rule-based model finds 52% of the malicious registrations
- The ML model finds 85% of the malicious registrations

Recall: How many malicious domains are selected? →

Data: Jan 2021 – Mar 2022

Use case: September 2022



Threshold for ML chosen such that number of stopped registrations ~ matches rule-based classifier

Around 88 per day

Use case: September 2022

Assumption: not verified = malicious

	Rules	ML
Precision	.49	.26
Recall	.90	.45
F1 score	.64	.33
F5 score	.87	.44

Assumption: not verified = benign

	Rules	ML
Precision	.01	.04
Recall	.13	.62
F1 score	.02	.08
F5 score	.08	.41

not verified => ignore

	Rules	ML
Precision	.01	.03
Recall	.10	.38
F1 score	.02	.06
F5 score	.08	.27

Conclusion

- Abusive registrations have distinct properties
    - Same/similar registration details
    - Fake contact data
    - Drop-catching domains
    - Similar domains

- Machine learning outperforms a rule-based system
- Ground truth is tricky
    - Bias towards rule-based system
    - Incompleteness of ground truth makes training & analysis hard

Goals of joint project

- Explore whether we can more effectively detect high-risk registrations through collaboration.

- Explore whether we can jointly develop a method to detect high-risk registrations.

- Explore whether we can develop a blueprint implementation and make that available to other registries.

Activities

- Share code ✓
- Learn from each others' assumptions and code
- Merge into single method,
  or extend individual methods
- Apply and evaluate each others' trained models
- Publish blueprint code

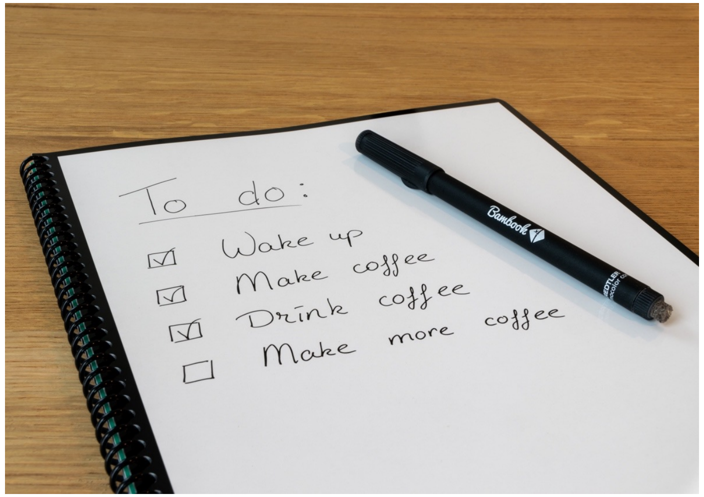


Photo by Thomas Borgmans on Unsplash

Commonalities (so far)

- Goal is proactively blocking high-risk registrations
- Verification process is as important as detection method
- Determining label definitions is a challenge (what is high risk?)
- Major policy component
- Coordination with stakeholders (support, policy, legal) takes time

Differences (so far)

DNS Belgium:

- Replacing an existing system
- Defer delegation automatically
- Focus on exploring features
- Focus on recall

SIDN:

- Starting with a clean state
- Review registrations manually
- Focus on mature implementation
- Focus on precision and interpretability

Conclusion

- Collaboration between .nl and .be this year
- Developing blueprint for detecting potentially malicious registrations
- Report back to CENTR community!

Q&A

thymen.wabeke@sidn.nl
thijs.vandenhout@sidn.nl
maarten.bosteels@dnsbelgium.be