Machine Learning for Defensive Cybersecurity

Azqa Nadeem

PhD candidate

Cyber Analytics Lab

DSC Delft

TU Delft

04 February 2021

Cyber Analytics Lab

> whoami

- Originally from Pakistan
- 3rd year PhD candidate
  - Sequential ML for network security
- Security lecturer
- Landscape photographer

> whoami



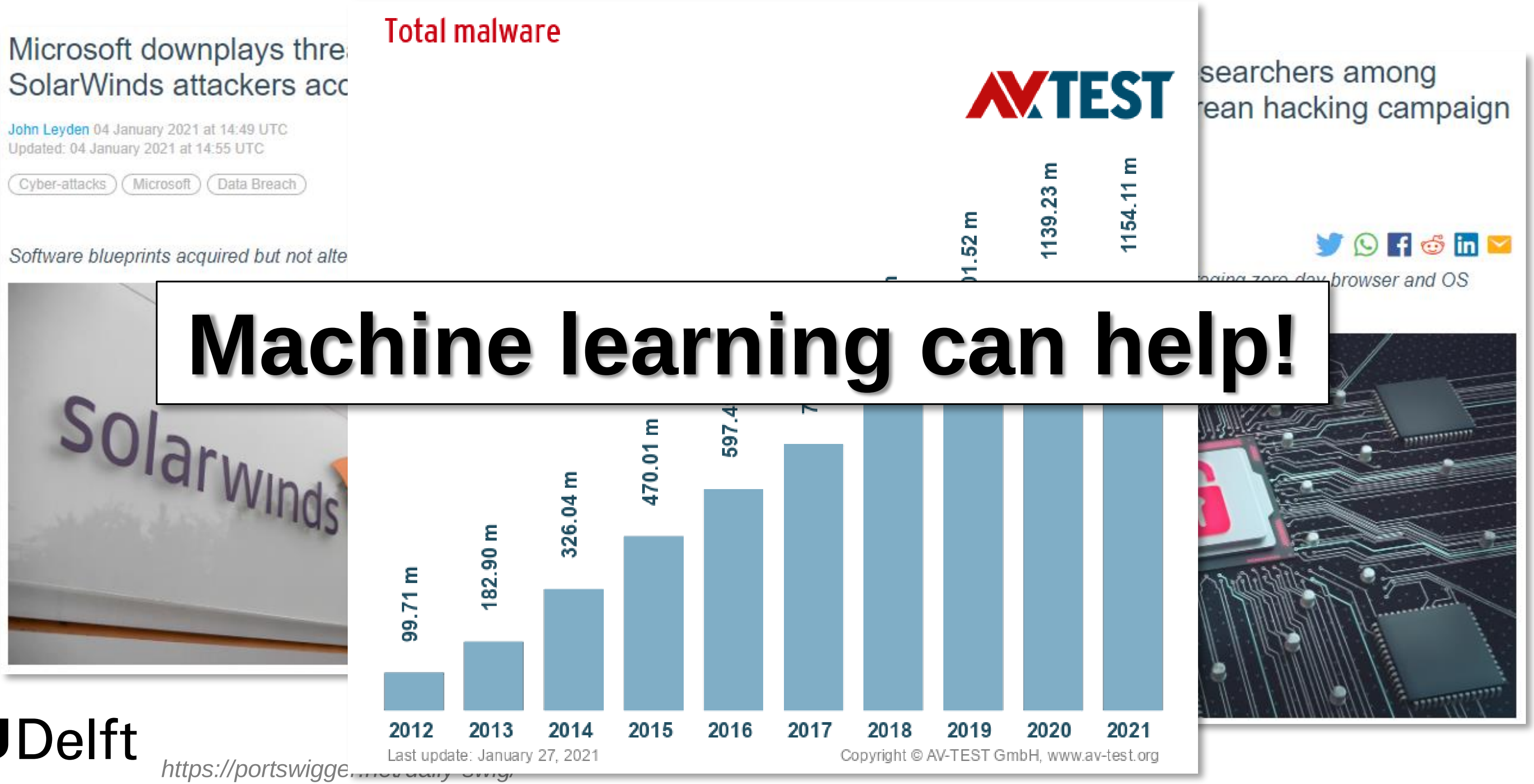






Azqa Nadeem

Current state of security

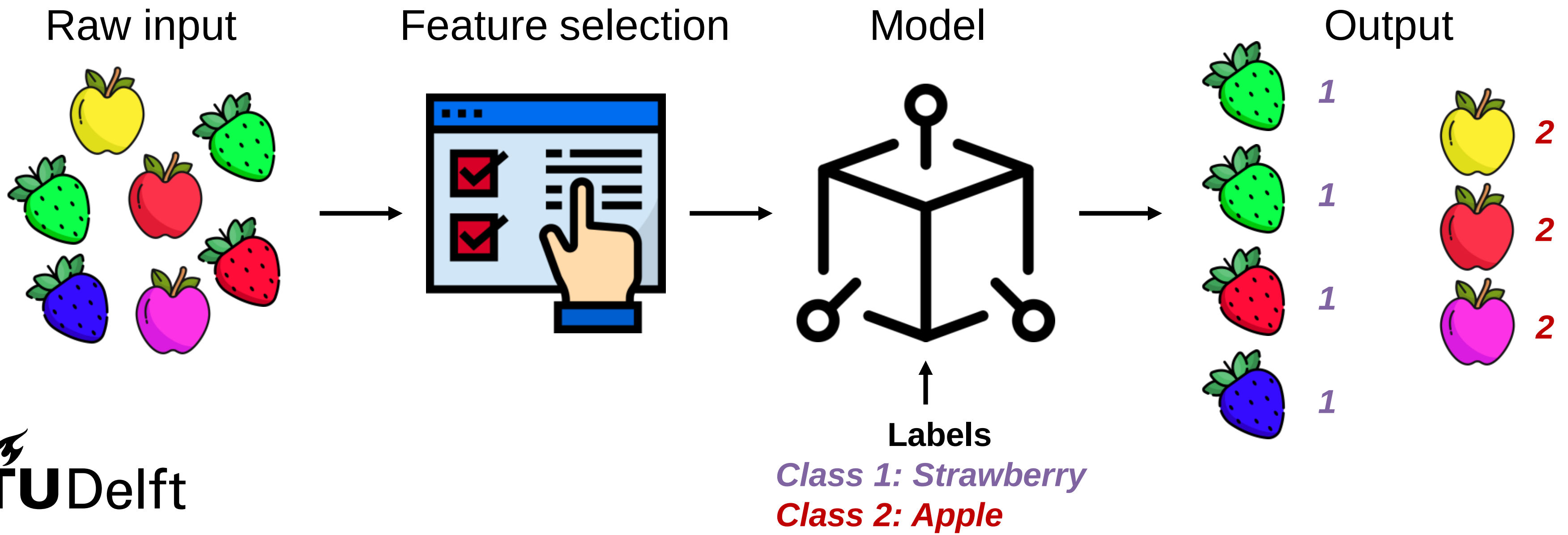


Machine learning can help!

https://portswigge...

Machine learning

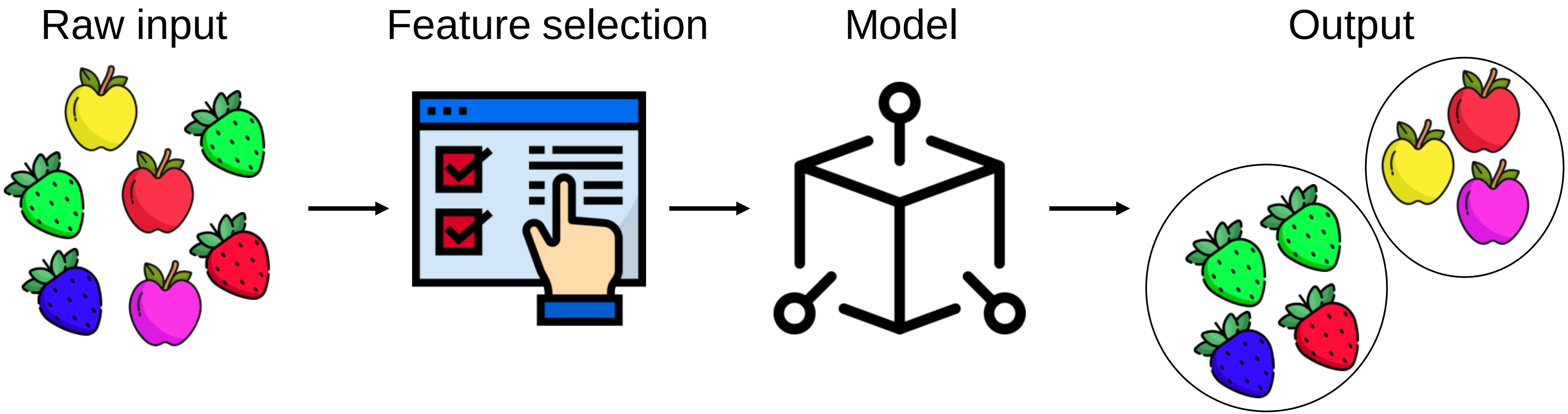
- Learn patterns from input data
- Under the hood: *Optimize an objective function*



TU Delft

Machine learning

- Learn patterns from input data
- Under the hood: *Optimize an objective function*



TU Delft

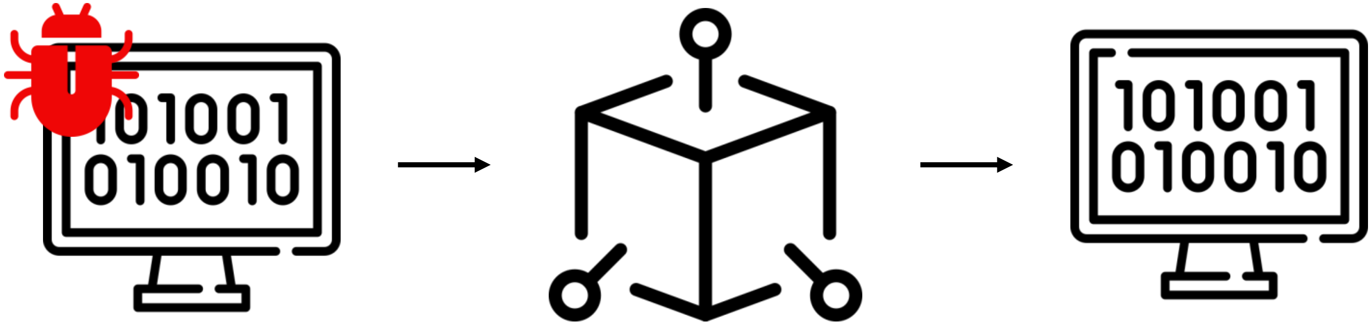
ML for defensive cybersecurity

- Spam detection
- Malware detection
- Detect and patch buggy code
- Detect real-time attacks
- Profile attacker behavior
- Anomaly detection
- Attacker modelling
  - APT modelling
- …

- *Offensive security applications*
  - *Crafting malware, hardware attacks, …*

ML for defensive cybersecurity

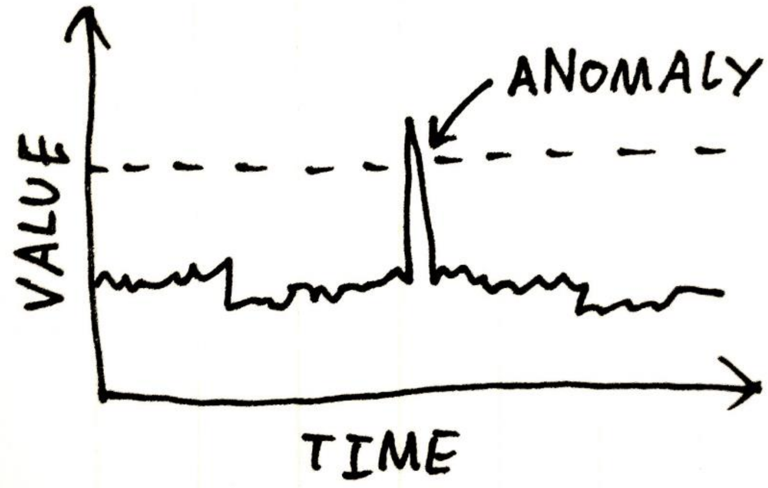
- Detect and patch buggy code



Tufano, Michele, et al. "An empirical study on learning bug-fixing patches in the wild via neural machine translation." ACM Transactions on Software Engineering and Methodology (TOSEM) 28.4, 2019.

ML for defensive cybersecurity

- Anomaly detection



Lin, Qin, et al. "TABOR: A graphical model-based approach for anomaly detection in industrial control systems." *Proceedings of the 2018 on Asia Conference on Computer and Communications Security. 2018.*
https://towardsdatascience.com/generating-critical-scenarios-using-anomaly-detection-f25e67e0553b

ML for defensive cybersecurity

- Malware detection → Predicting impending exposure

Sharif, Mahmood, et al. "Predicting impending exposure to malicious content from user behavior." Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. 2018.

ML for defensive cybersecurity

- Malware detection → Capability assessment

Behavior profile

- Connects with C&C
- Opens backdoors
- Persistent

vs.

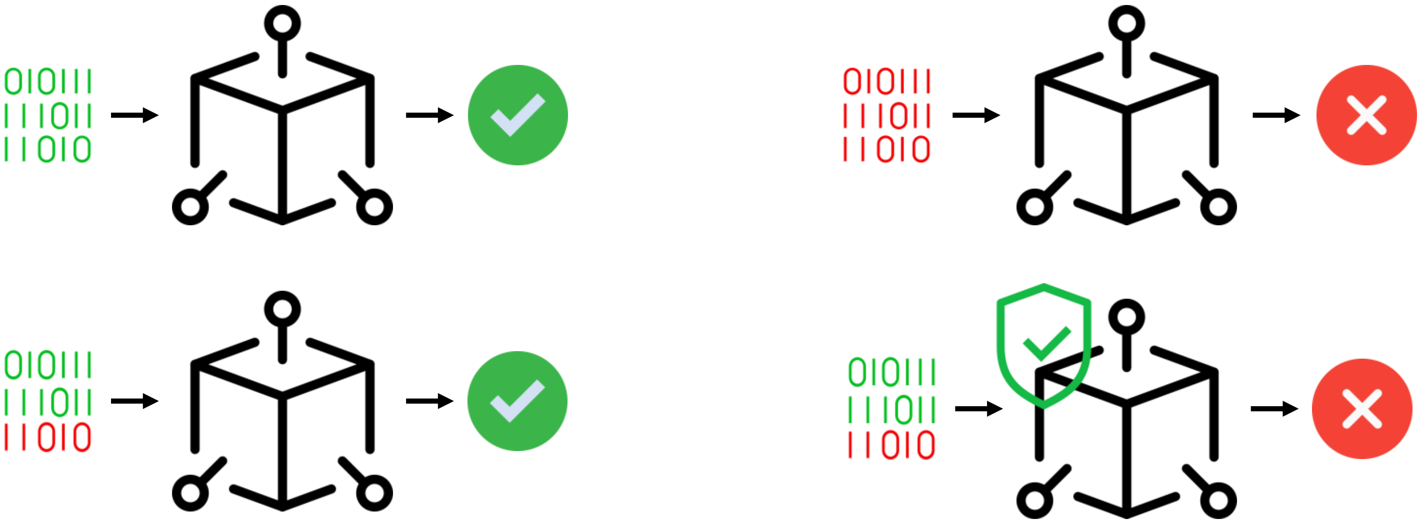
Label

ZeuS

Nadeem, Azqa, et al. "Beyond Labeling: Using Clustering to Build Network Behavioral Profiles of Malware Families." Malware Analysis Using Artificial Intelligence and Deep Learning. Springer, Cham, 2021.

ML for defensive cybersecurity

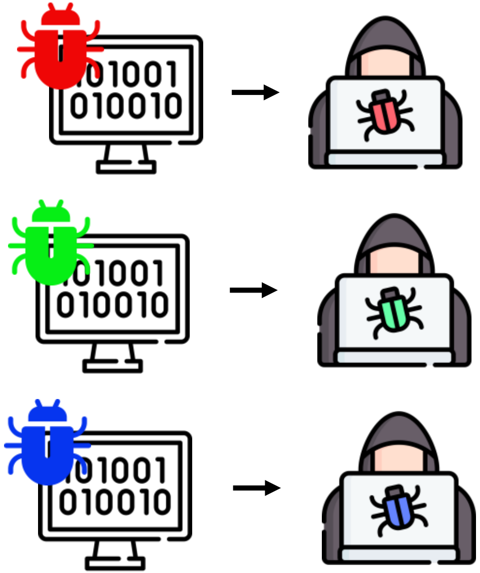
- Malware detection → Adversarial ML



Verwer, Sicco, et al. "The Robust Malware Detection Challenge and Greedy Random Accelerated Multi-Bit Search." *Proceedings of the 13th ACM Workshop on Artificial Intelligence and Security.* 2020.

ML for defensive cybersecurity

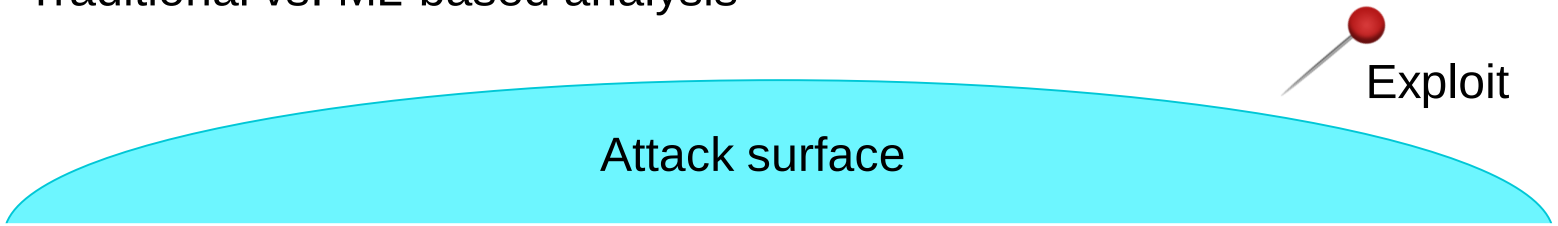
- Malware detection → Author attribution



Murenin, Ivan, et al. "Explaining Android Application Authorship Attribution Based on Source Code Analysis." Internet of Things, Smart Spaces, and Next Generation Networks and Systems. Springer, Cham, 2020.

Industry perspective

- Divide between academia & industry
- ML's slow adaptation
  - Traditional vs. ML-based analysis



ML is not a silver bullet

- Cannot blindly apply ML to Security
  - Address unique problems

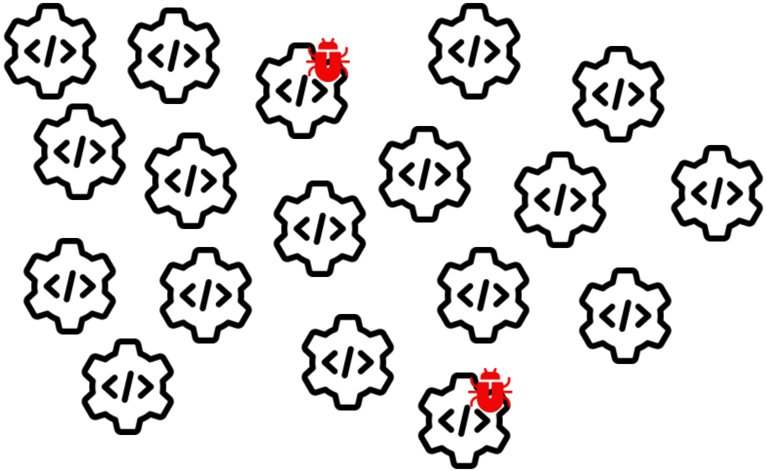
- Do not throw data in black-box
  - Ethical considerations

(Caution!) More goodware than malware [1/4]

- Security data has class imbalance

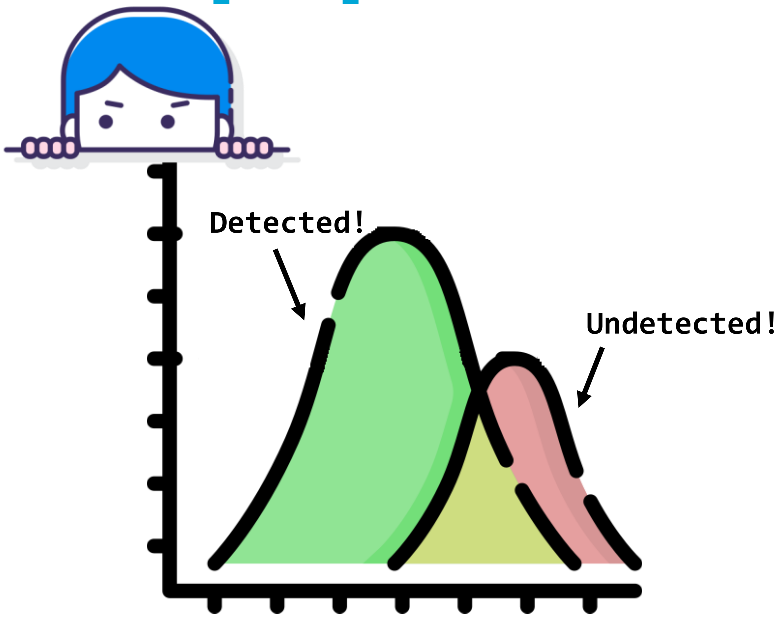
- Unrealistic class distribution
  - Bias in data → bias in models

- Use real class distribution
- Use imbalance-aware algorithms



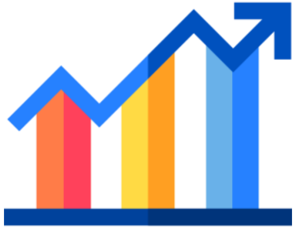
(Caution!) Landscape is adversarial [2/4]

- Attackers hide, malware evades detection
- ML cannot detect all evasion attempts!
- Representative dataset is required!
- ML can adapt to changing landscape
  - Trigger re-learning



(Caution!) Know what to evaluate [3/4]

- Be mindful of evaluation metrics
  - Precision, Recall, AUC, F1 score …
  - Accuracy in imbalanced datasets



- Performance metrics ≠ improved security

- Better understanding fosters better models
  - Prediction vs. understandability

(Caution!) Know the limitations of ML [4/4]

- Can find patterns faster than humans
  - But is also really stupid
- Cannot replace human intelligence
  - Trade-off between automation and explainability
- Build human-in-the-loop ML pipelines

https://hackernoon.com/human-intelligence-or-artificial-intelligence-we-need-both-if7w32b2

Take-aways

- ML enables human analysts to do complex tasks
  - A powerful technology for defensive security
  - But cannot blindly apply it
- ML used for both defense and offense
  - Performance metrics ≠ security
  - Robust classifiers required
- ML is not a silver bullet for all security problems
  - Explainable and Human-in-the-loop ML is paramount

Thank you!

azqa.nadeem@tudelft.nl @azqa_nadeem https://cyber-analytics.nl/

TU Delft