Demystifying Artificial Intelligence and Machine Learning

John Strand, IANS Faculty

Agenda

- AI and ML – An Overview
- ML Uses in Security
- Security Analytics – A Set of Products
- Action Plan
- What Will the Future Bring?

AI & ML

"Everyone calls their stuff 'machine learning' or even better 'artificial intelligence' - It's not cool to use statistics!"

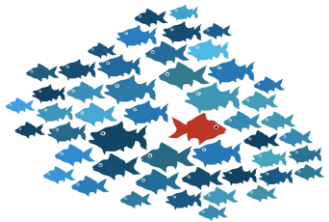
AI and ML are promising approaches to solve **some** security problems:

- These are algorithms, not products
- Expert knowledge is more important than algorithms
- Don't start your own ML projects unless you have the right data and skills
- *Buy products that address your use-cases*

AI & ML - An Overview

Machine Learning & Artificial Intelligence

- **Machine Learning (ML)**
  - Learns from training data to classify data (e.g., SPAM or malware classification)
- **Anomaly Detection (Outlier Detection)**
  - Can be done with ML but simple statistics often work much better
  - Statistical outliers are hardly ever security relevant
  - 2 decades of anomaly detection research in security!

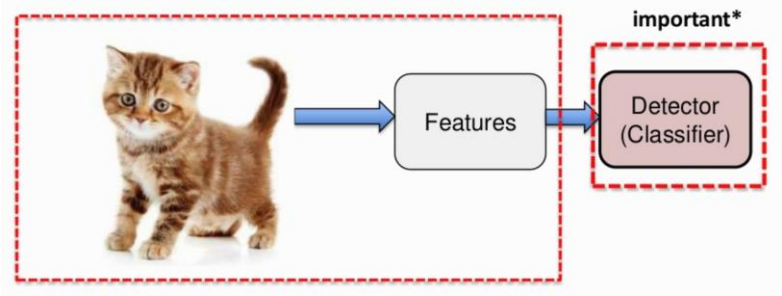


- **Deep learning**
  - Is just another ML algorithm - significantly improved results for classification problems
  - Basically eliminates the feature engineering step
- **Artificial Intelligence (AI)**
  - *"A program that doesn't simply classify or compute model parameters, but comes up with **novel knowledge** that a security analyst finds insightful."*

IANS

ML & Deep Learning

- Any Machine Learning:
  - Is used to **classify** data
  - Needs a lot of well-labeled **training** data
  - Good for malware / SPAM identification



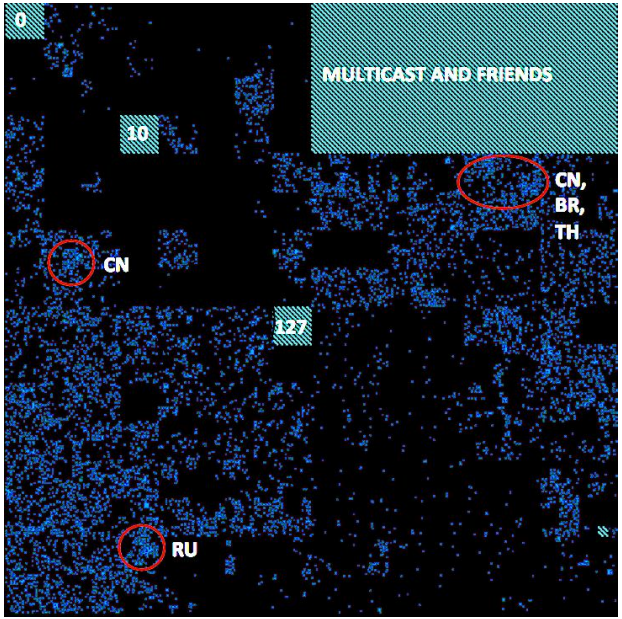
- Traditional Machine Learning:
  - Features / Attributes identified by experts
  - Similarity and correlation determined by statistical evaluation and expert knowledge

- Deep Learning:
  - Automatically **learns features** from data
  - Eliminates the bias introduced by the human identification of features
  - Smaller model, faster and more accurate
  - **Lacks explain-ability**

IANS

ML Uses in Security

- Malware detection and classification
- SPAM identification
- Firewall data analysis to identify likely attackers
- DNS analytics:
  - Co-occurrence, domain name classification
  - DNS lookup analysis (frequency)
- Threat intelligence feed analysis:
  - IOC prioritization, de-duplication, campaign association, removing false positives
- URL analytics:
  - Identify malicious URLs
  - Turns out, you have to analyze the content of the website behind the URL as well



www.mlsecproject.org

IANS

Security Analytics - A Set of Products

Attackers are using ‘allowed’ channels and mask in benign looking activity that traditional security tools cannot detect.

User and Entity Behavior Analytics (UEBA)

- Identify anomalies based on user and/or machine behavior.
- Most vendors don’t use real machine learning, don’t fall for snake oil – ask for real-world proof
- Two groups of products: based on logs or based on network traffic

Automation & Orchestration

- Sit on top of SIEM (and some other data) to close the loop of a) prioritizing important attacks and b) automating response.

Hunting

- Enable senior security analysts to explore data within a SIEM or big data store to find environment specific attacks and breaches.

All these products are really features of a larger platform:

- They should all be under one single product
- If they are sold as individual products, make sure they interoperate well. Where is the data stored? etc.

IANS

What Now?

Action Plan

- Define your **use-cases** first - understand where you want and should use ML
- Make sure you have the right **data and context**
- **Understand your environment** inside out!
- Buy **products** for your most pressing problems. Make sure they solve them cost-effectively!
- Don’t ever have an “AI project”

Practical Considerations Buying Analytics Products

- Does the solution really detect **behavioral anomalies**?
- Does the solution **integrate** with the rest of your infrastructure (e.g., SIEM)?
- **How long** does it take to begin recognizing suspicious patterns? How long does it take to establish a baseline?
- How does the solution adapt to completely **novel attacks**?
- Ask for **results** that have been seen in actual customer environments
- Do a **PoC on your network** to learn:
  - How hard it is to **install** the product and how much time does it take to **tune?**
  - How much time it will take on **ongoing** maintenance?
  - What does it actually **detect** in your environment?

IANS

What Will the Future Bring?

- More bad marketing calling ‘analytics’ and ‘statistics’ ML and AI
- Algorithms will advance, but we won’t hit real AI anytime soon
- Consolidations in the ‘ Security Analytics’ product space
- More and better approaches to model expert knowledge
- Data lakes will eventually become a reality - analytics will have to run on top of that
- Better, automated asset inventory
- Data sharing will become more and more important (TI, models, etc.)
  - Threat Intelligence will morph more and more into real-time data sharing between trusted entities
- Talent gap will keep widening - How do you staff your projects?

IANS

Summary

- We don’t have **artificial intelligence** (yet)
- **Machine learning** is an algorithm not a product
- Algorithms are getting ‘smarter’, but **experts** are more important
- Invest in **people** who know security (and have experience)
- **Understand** your environments, applications, devices
- Focus on advancing **insights**

IANS

BlackHat Workshop

Applied Machine Learning
for
Identity and Access Management

ML | AI | IAM

August 4,5 & August 6,7 - Las Vegas, USA

http://secviz.org

IANS

Questions?

info@iansresearch.com