OWASP

Open Web Application
Security Project

Machine Learning for Application Security Professionals

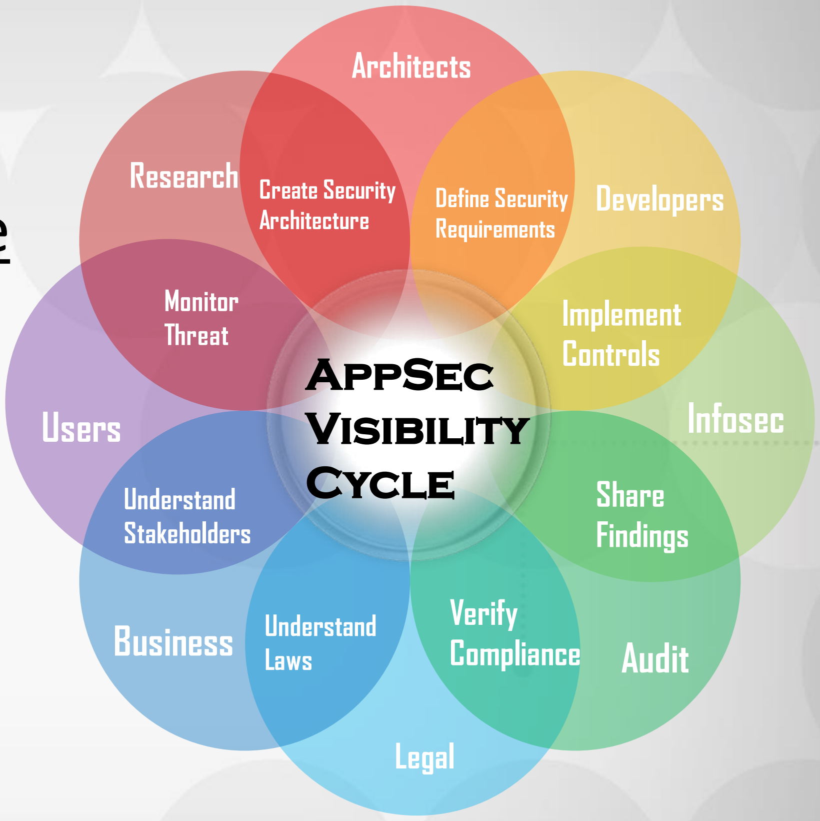
Talal Albacha

@talal_basha1982

Talal.basha@owasp.org

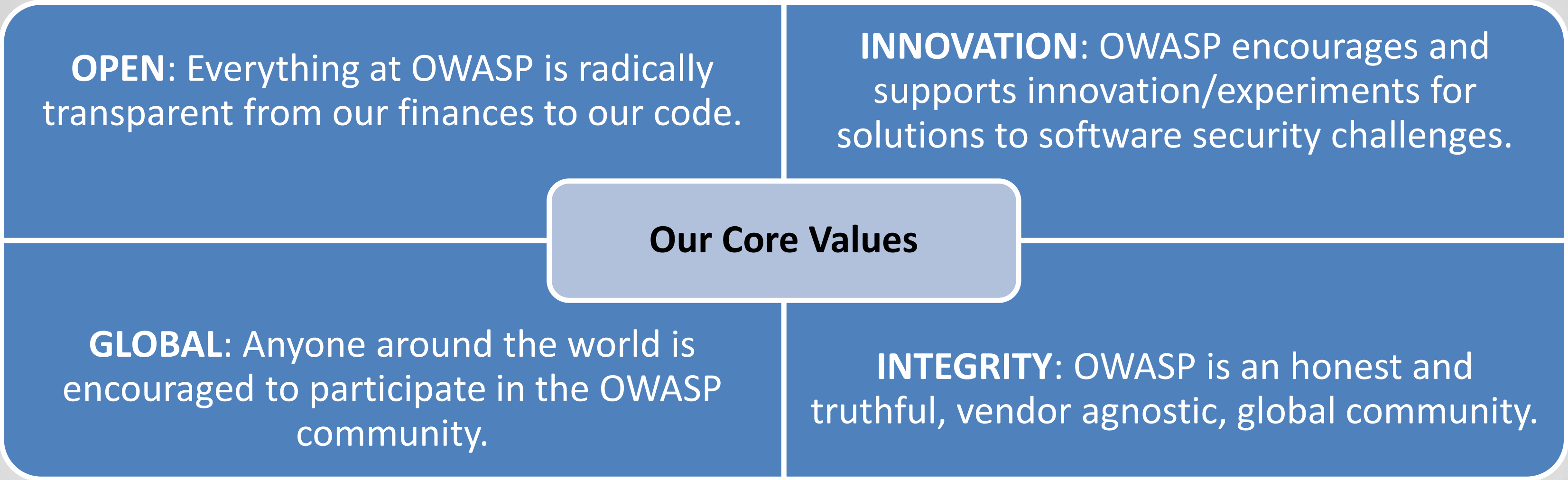
Mission

- Our mission is to make software security visible, so that individuals and organizations worldwide can make informed decisions about true software security risks



Our Purpose & Our Core Values

Our Purpose: The OWASP Foundation will be the thriving global community that drives visibility and evolution in the safety and security of the world's software.



Our Core Values

OPEN: Everything at OWASP is radically transparent from our finances to our code.

INNOVATION: OWASP encourages and supports innovation/experiments for solutions to software security challenges.

GLOBAL: Anyone around the world is encouraged to participate in the OWASP community.

INTEGRITY: OWASP is an honest and truthful, vendor agnostic, global community.

Flagship Projects [edit source]

The OWASP Flagship designation is given to projects that have demonstrated strategic value to OWASP and application security as a whole. After a major review process [More info here] the following projects are considered to be flagship candidate projects. These project have been evaluated more deeply to confirm their flagship status:

Tools [Health Check January 2017] [edit source]

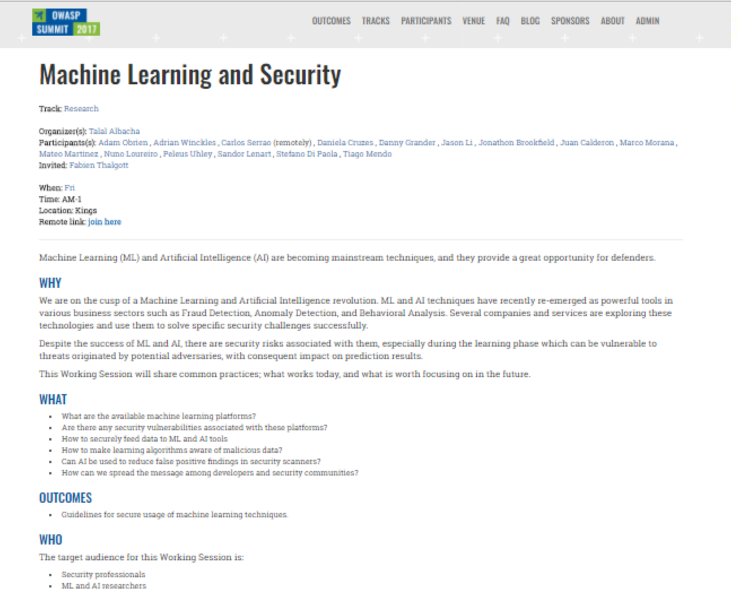
- OWASP Zed Attack Proxy👍
- OWASP Web Testing Environment Project👍
- OWASP OWTF👍
- OWASP Dependency Check👍
- OWASP Security Shepherd👍

Code [Health Check January 2017] [edit source]

- OWASP ModSecurity Core Rule Set Project👍
- OWASP CSRFGuard Project👍
- OWASP AppSensor Project👍

Documentation[Health Check January 2017] [edit source]

- OWASP Application Security Verification Standard Project👍
- OWASP Software Assurance Maturity Model (SAMM)👍
- OWASP AppSensor Project👍
- OWASP Top Ten Project👍
- OWASP Testing Project👍



OUTCOMES TRACKS PARTICIPANTS VENUE FAQ BLOG SPONSORS ABOUT ADMIN

Machine Learning and Security

Track: Research

Organizer(s): Talal Albacha
Participants(s): Adam Obrien , Adrian Winckles , Carlos Serrao (remotely) , Daniela Cruzes , Danny Grander , Jason Li , Jonathon Brookfield , Juan Calderon , Marco Morana , Mateo Martinez , Nuno Loureiro , Peleus Uhley , Sandor Lenart , Stefano Di Paola , Tiago Mendo
Invited: Fabien Thalgott

When: Fri
Time: AM-1
Location: Kings
Remote link: join here

Machine Learning (ML) and Artificial Intelligence (AI) are becoming mainstream techniques, and they provide a great opportunity for defenders.

WHY

We are on the cusp of a Machine Learning and Artificial Intelligence revolution. ML and AI techniques have recently re-emerged as powerful tools in various business sectors such as Fraud Detection, Anomaly Detection, and Behavioral Analysis. Several companies and services are exploring these technologies and use them to solve specific security challenges successfully.

Despite the success of ML and AI, there are security risks associated with them, especially during the learning phase which can be vulnerable to threats originated by potential adversaries, with consequent impact on prediction results.

This Working Session will share common practices; what works today, and what is worth focusing on in the future.

WHAT

- What are the available machine learning platforms?
- Are there any security vulnerabilities associated with these platforms?
- How to securely feed data to ML and AI tools
- How to make learning algorithms aware of malicious data?
- Can AI be used to reduce false positive findings in security scanners?
- How can we spread the message among developers and security communities?

OUTCOMES

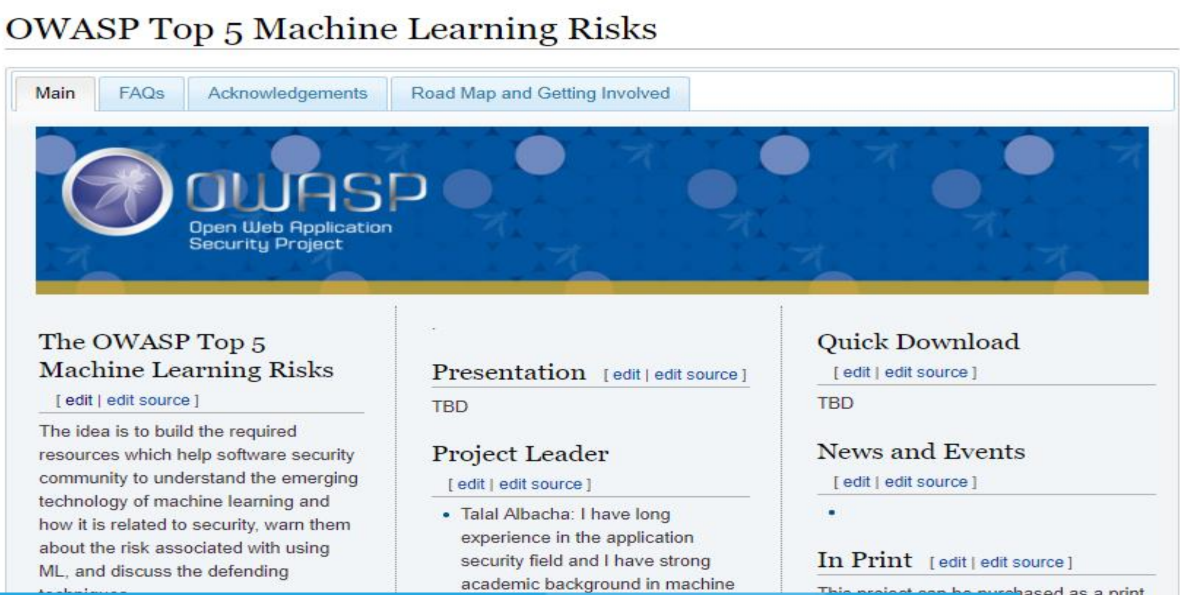
- Guidelines for secure usage of machine learning techniques.

WHO

The target audience for this Working Session is:

- Security professionals
- ML and AI researchers
- Devops
- SOC teams

WORKING MATERIALS



OWASP Top 5 Machine Learning Risks

Main | FAQs | Acknowledgements | Road Map and Getting Involved

OWASP Open Web Application Security Project

The OWASP Top 5 Machine Learning Risks

[edit | edit source]

The idea is to build the required resources which help software security community to understand the emerging technology of machine learning and how it is related to security, warn them about the risk associated with using ML, and discuss the defending techniques

Presentation [edit | edit source]

TBD

Project Leader

[edit | edit source]

- Talal Albacha: I have long experience in the application security field and I have strong academic background in machine

Quick Download

[edit | edit source]

TBD

News and Events

[edit | edit source]

-

In Print [edit | edit source]

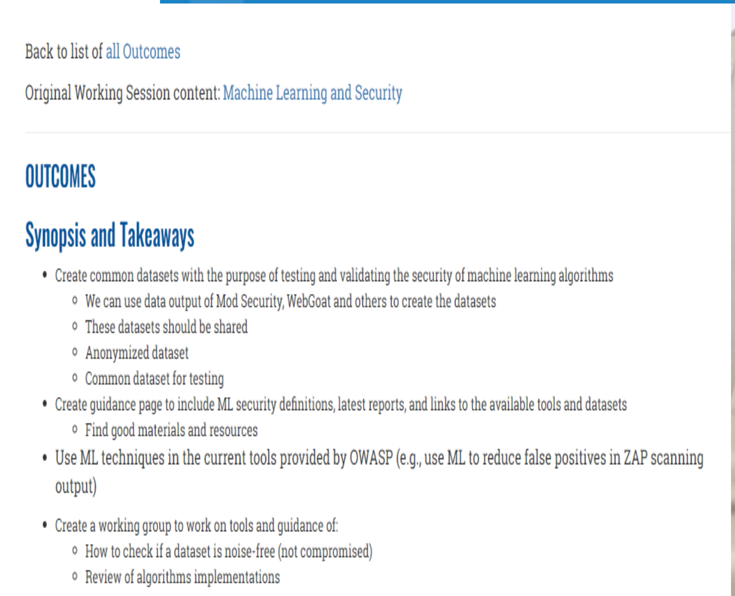
This project can be purchased as a print
lu.com

edit | edit source]

ders

OWASP SUMMIT 2017

Machine L



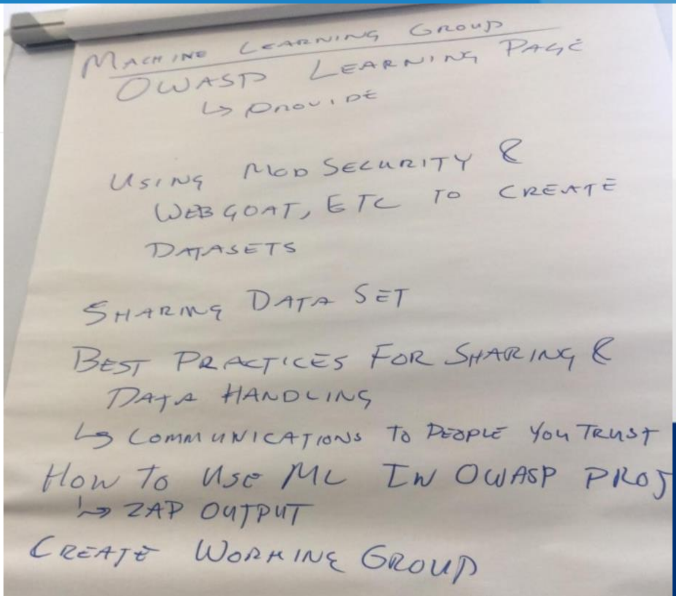
Back to list of all Outcomes

Original Working Session content: Machine Learning and Security

OUTCOMES

Synopsis and Takeaways

- Create common datasets with the purpose of testing and validating the security of machine learning algorithms
  - We can use data output of Mod Security, WebGoat and others to create the datasets
  - These datasets should be shared
  - Anonymized dataset
  - Common dataset for testing
- Create guidance page to include ML security definitions, latest reports, and links to the available tools and datasets
  - Find good materials and resources
- Use ML techniques in the current tools provided by OWASP (e.g., use ML to reduce false positives in ZAP scanning output)
- Create a working group to work on tools and guidance of:
  - How to check if a dataset is noise-free (not compromised)
  - Review of algorithms implementations



MACHINE LEARNING GROUP
OWASP LEARNING PAGE
↳ PROVIDE

USING MOD SECURITY &
WEB GOAT, ETC TO CREATE
DATASETS

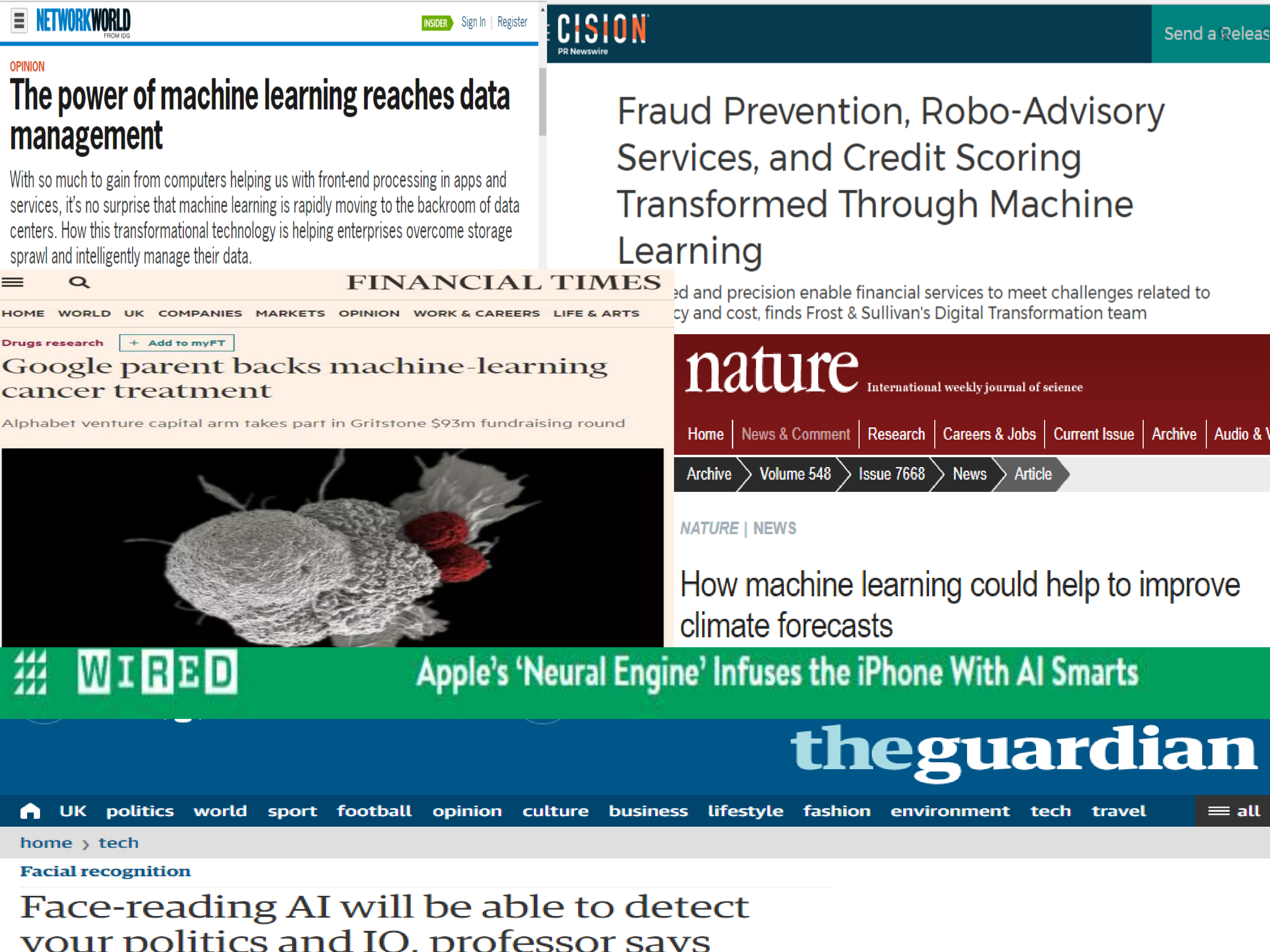
SHARING DATA SET

BEST PRACTICES FOR SHARING &
DATA HANDLING
↳ COMMUNICATIONS TO PEOPLE YOU TRUST

HOW TO USE ML IN OWASP PROJ
↳ ZAP OUTPUT

CREATE WORKING GROUP

OWASP
Open Web Application
Security Project



OPINION

The power of machine learning reaches data management

With so much to gain from computers helping us with front-end processing in apps and services, it's no surprise that machine learning is rapidly moving to the backroom of data centers. How this transformational technology is helping enterprises overcome storage sprawl and intelligently manage their data.

Fraud Prevention, Robo-Advisory Services, and Credit Scoring Transformed Through Machine Learning

ed and precision enable financial services to meet challenges related to
cy and cost, finds Frost & Sullivan's Digital Transformation team

Drugs research

Google parent backs machine-learning cancer treatment

Alphabet venture capital arm takes part in Gritstone $93m fundraising round

NATURE | NEWS

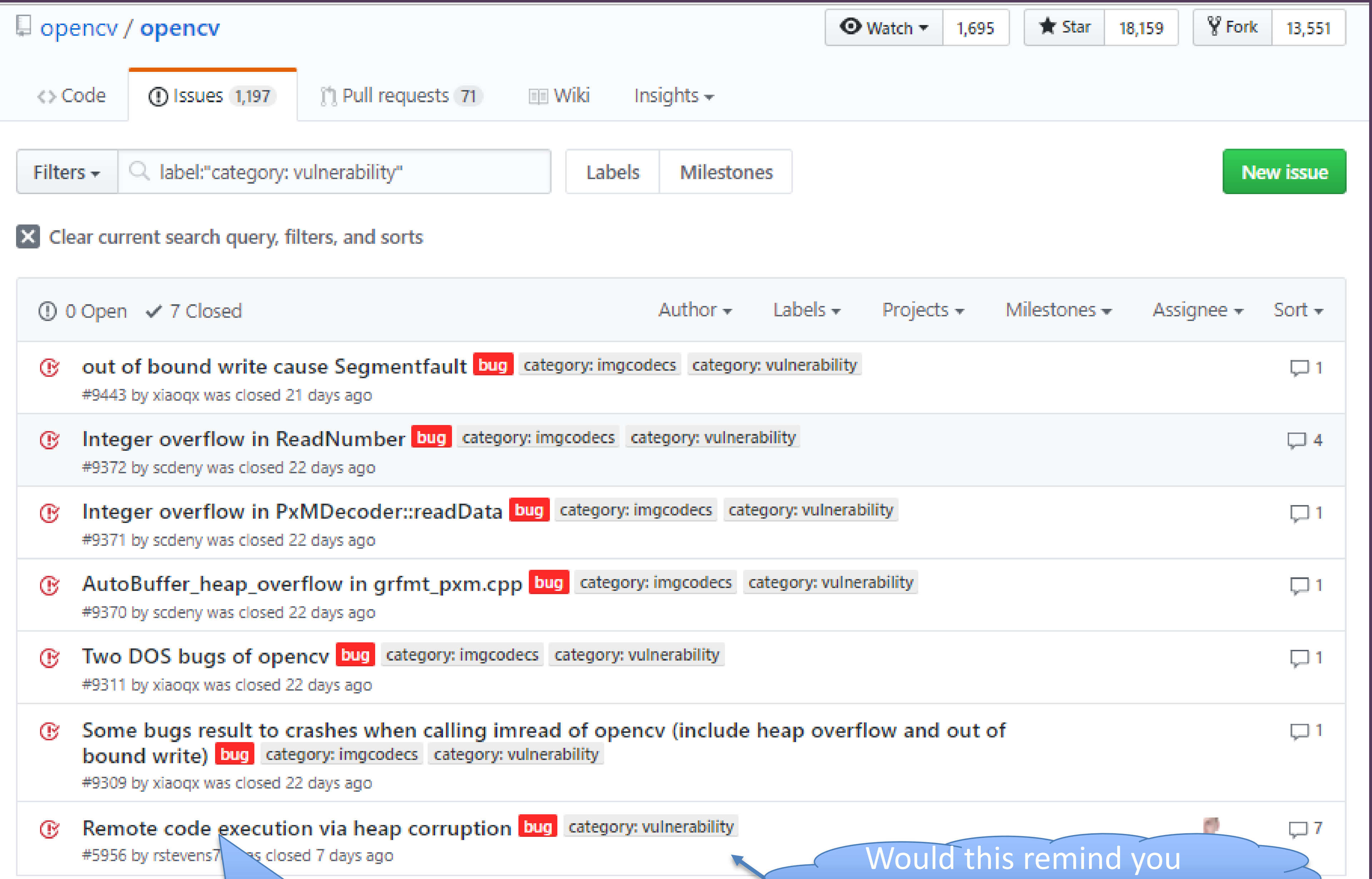
How machine learning could help to improve climate forecasts

Apple's 'Neural Engine' Infuses the iPhone With AI Smarts

Facial recognition

Face-reading AI will be able to detect your politics and IQ, professor says

Definitions



label:"category: vulnerability"

Clear current search query, filters, and sorts

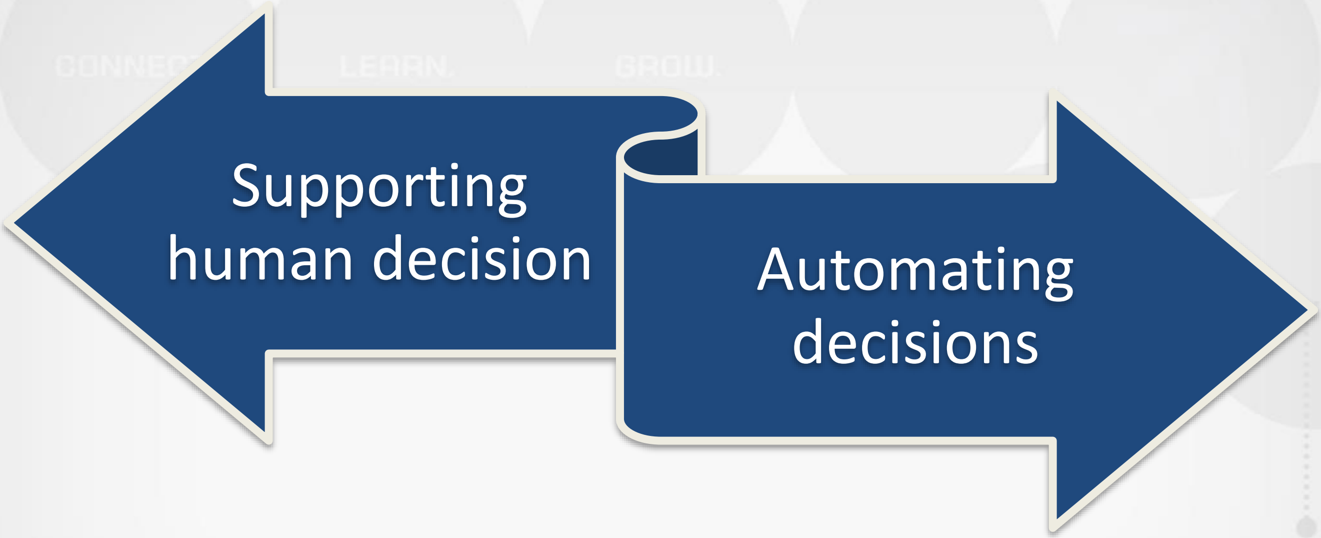
0 Open ✓ 7 Closed

- out of bound write cause Segmentfault `bug` `category: imgcodecs` `category: vulnerability` — 1
  #9443 by xiaoqx was closed 21 days ago
- Integer overflow in ReadNumber `bug` `category: imgcodecs` `category: vulnerability` — 4
  #9372 by scdeny was closed 22 days ago
- Integer overflow in PxMDecoder::readData `bug` `category: imgcodecs` `category: vulnerability` — 1
  #9371 by scdeny was closed 22 days ago
- AutoBuffer_heap_overflow in grfmt_pxm.cpp `bug` `category: imgcodecs` `category: vulnerability` — 1
  #9370 by scdeny was closed 22 days ago
- Two DOS bugs of opencv `bug` `category: imgcodecs` `category: vulnerability` — 1
  #9311 by xiaoqx was closed 22 days ago
- Some bugs result to crashes when calling imread of opencv (include heap overflow and out of bound write) `bug` `category: imgcodecs` `category: vulnerability` — 1
  #9309 by xiaoqx was closed 22 days ago
- Remote code execution via heap corruption `bug` `category: vulnerability` — 7
  #5956 by rstevens7 was closed 7 days ago

CVE-2016-1516

Would this remind you of any recent RCE ..

$$CL = (CI/CD)^2$$



Supporting human decision

Automating decisions

How did you reach to this decision?

We need root cause analysis

We need Audit Trail

By Experience ;)

$$0.-30.1996\,i)\,\zeta^6+(0.+9.81486\,i)\,\zeta^5\sqrt{\zeta^2-(0.+640.701\,i)}\;+$$

$$(0.+20.3847\,i)\,\zeta^5\sqrt{\zeta^2-(0.+1127.53\,i)}\;+(0.+20.3847\,i)\,\zeta^4\left(\zeta^2-(0.+640.701\,i)\right)+$$

$$(0.+9.81486\,i)\,\zeta^4\left(\zeta^2-(0.+1127.53\,i)\right)-(0.+9.81486\,i)\,\zeta^3\sqrt{\zeta^2-(0.+640.701\,i)}\;\left(\zeta^2-(0.+1127.53\,i)\right)-$$

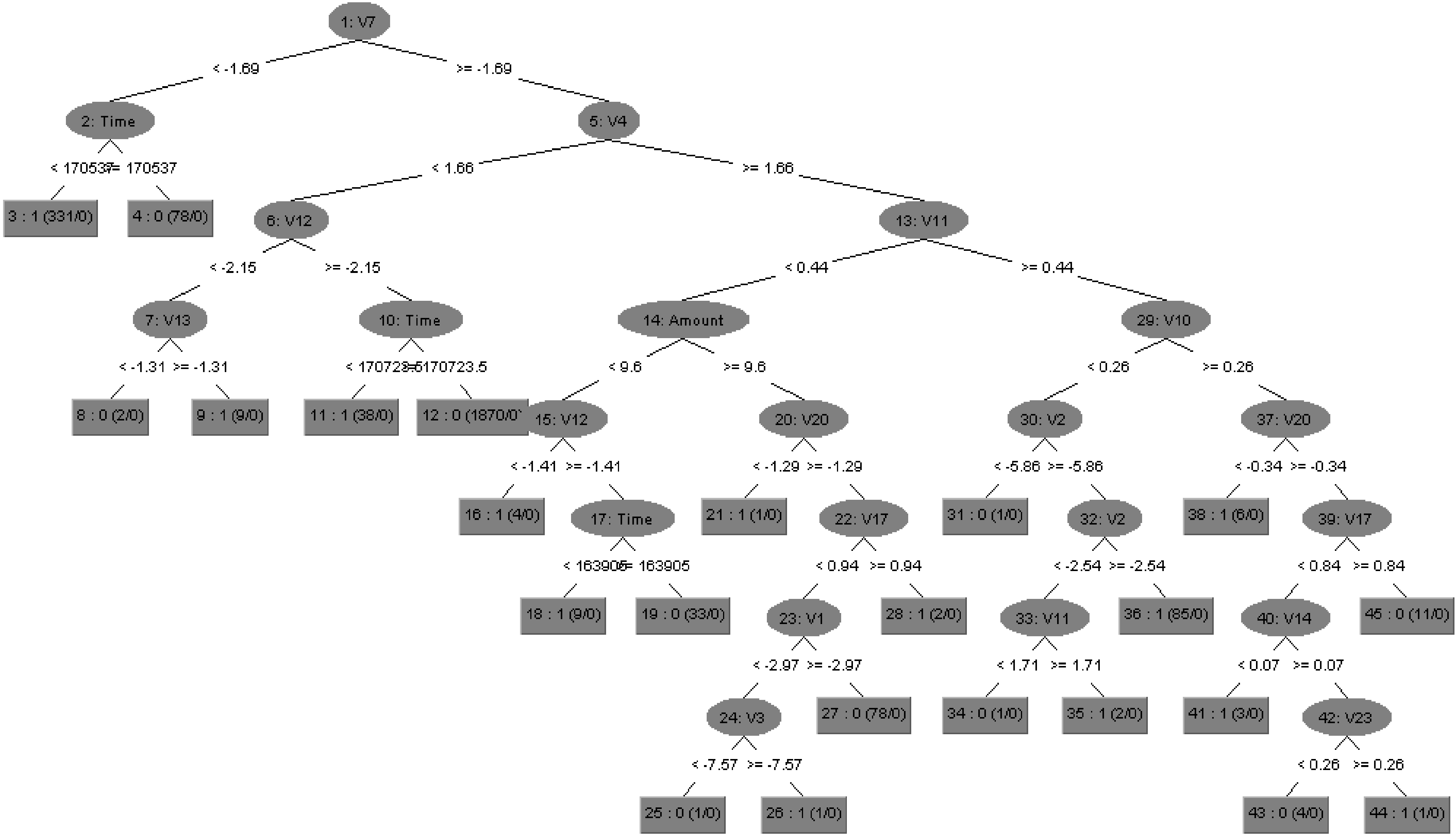
$$(0.+20.3847\,i)\,\zeta^3\left(\zeta^2-(0.+640.701\,i)\right)\sqrt{\zeta^2-(0.+1127.53\,i)}\;+$$

$$408.549\left((0.-0.630868\,i)\,\zeta^3-(0.+14.0909\,i)\,\zeta^2\sqrt{\zeta^2-(0.+640.701\,i)}\;-\right.$$

$$(0.+6.45738\,i)\,\zeta^2\sqrt{\zeta^2-(0.+1127.53\,i)}\;-(0.+1.21667\,i)\,\zeta\left(\zeta^2-(0.+640.701\,i)\right)+$$

$$(0.+0.585806\,i)\,\zeta\left(\zeta^2-(0.+1127.53\,i)\right)+(0.+7.08825\,i)\sqrt{\zeta^2-(0.+640.701\,i)}\;\left(\zeta^2-(0.+1127.53\,i)\right)$$

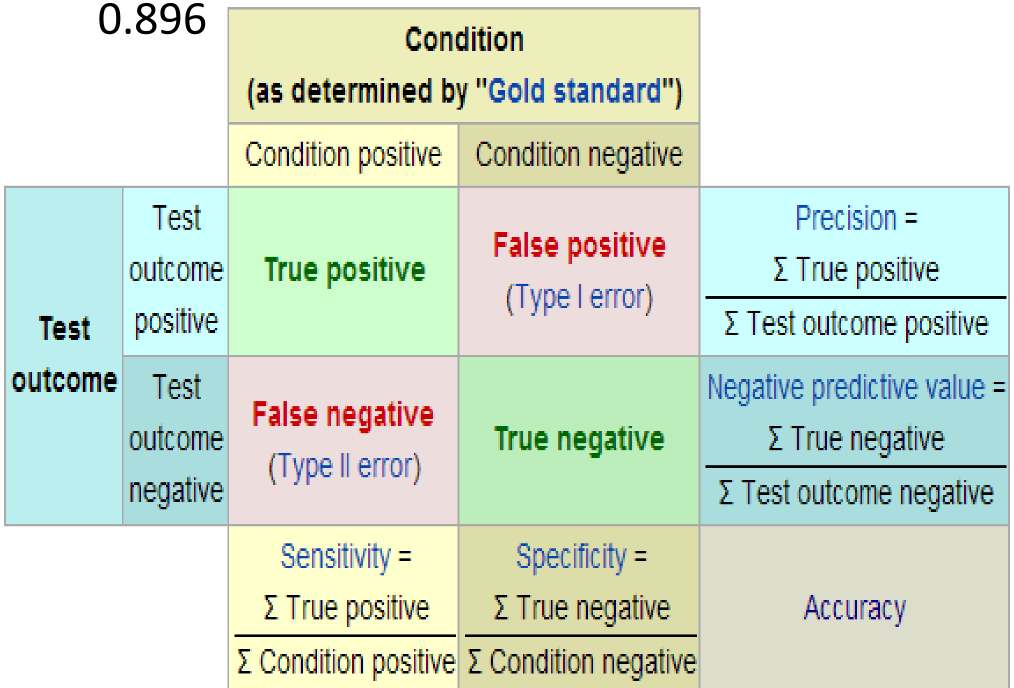
$$\zeta^2-(0.+640.701\,i)$$


1: V7
< -1.69
>= -1.69
2: Time
5: V4
< 170537>= 170537
< 1.66
>= 1.66
3 : 1 (331/0)
4 : 0 (78/0)
6: V12
13: V11
< -2.15
>= -2.15
< 0.44
>= 0.44
7: V13
10: Time
14: Amount
29: V10
< -1.31
>= -1.31
< 170723.5>=170723.5
< 9.6
>= 9.6
< 0.26
>= 0.26
8 : 0 (2/0)
9 : 1 (9/0)
11 : 1 (38/0)
12 : 0 (1870/0
15: V12
20: V20
30: V2
37: V20
< -1.41
>= -1.41
< -1.29
>= -1.29
< -5.86
>= -5.86
< -0.34
>= -0.34
16 : 1 (4/0)
17: Time
21 : 1 (1/0)
22: V17
31 : 0 (1/0)
32: V2
38 : 1 (6/0)
39: V17
< 163905>= 163905
< 0.94
>= 0.94
< -2.54
>= -2.54
< 0.84
>= 0.84
18 : 1 (9/0)
19 : 0 (33/0)
23: V1
28 : 1 (2/0)
33: V11
36 : 1 (85/0)
40: V14
45 : 0 (11/0)
< -2.97
>= -2.97
< 1.71
>= 1.71
< 0.07
>= 0.07
24: V3
27 : 0 (78/0)
34 : 0 (1/0)
35 : 1 (2/0)
41 : 1 (3/0)
42: V23
< -7.57
>= -7.57
< 0.26
>= 0.26
25 : 0 (1/0)
26 : 1 (1/0)
43 : 0 (4/0)
44 : 1 (1/0)

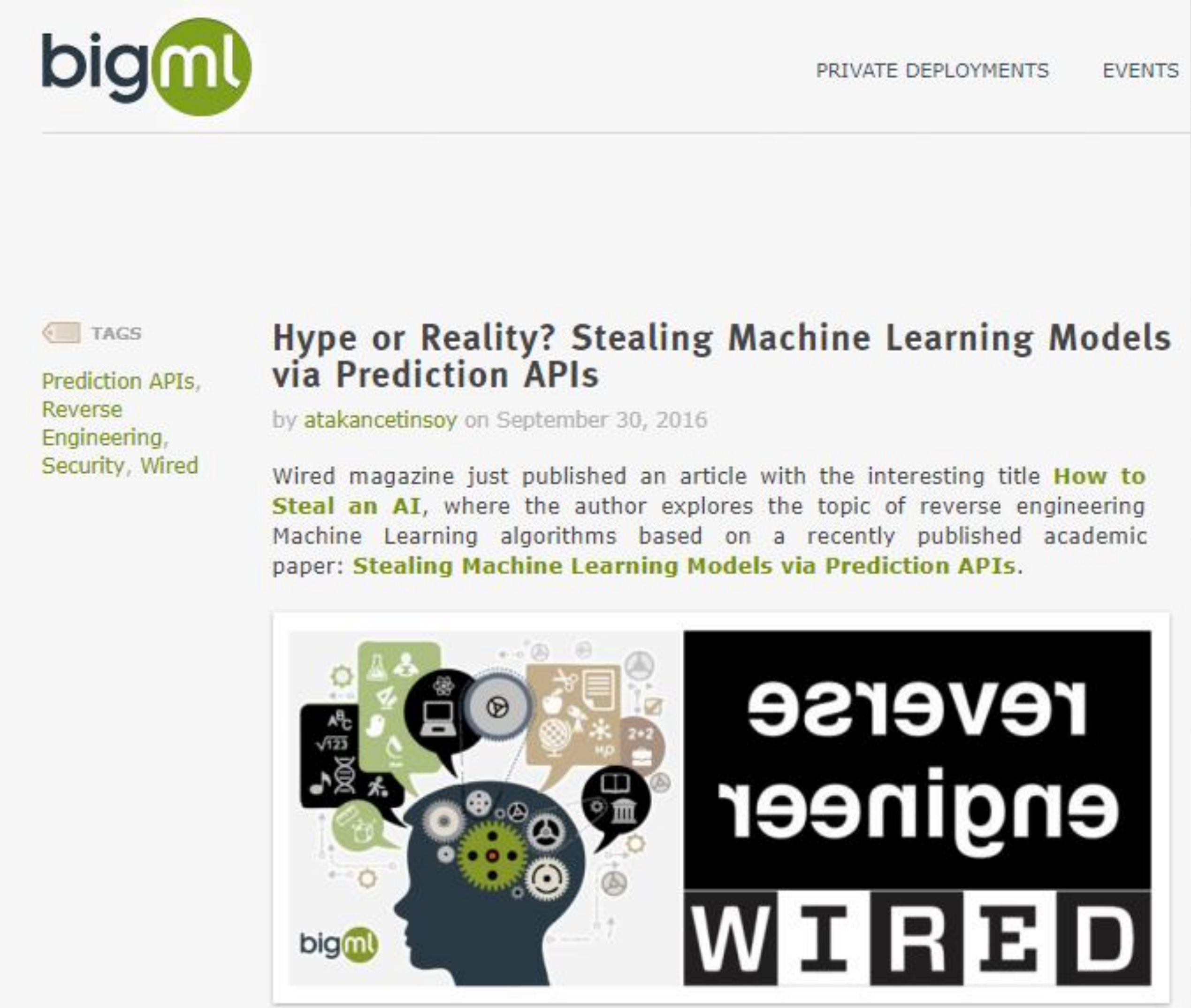
Counting false positives only is not accurate

```
Correctly Classified Instances        83      %
Incorrectly Classified Instances      17      %
=== Detailed Accuracy By Class ===

                    TP Rate  FP Rate  ROC Area  PRC Area  Class
                    0.909    0.324    0.897     0.950     fraud
                    0.676    0.091    0.897     0.790     not_fraud
Weighted Avg.       0.830    0.244    0.897     0.896
```



		Condition **(as determined by "Gold standard")**		
		Condition positive	Condition negative	
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)	Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Test outcome positive}}$
	Test outcome negative	**False negative** (Type II error)	**True negative**	Negative predictive value = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Test outcome negative}}$
		Sensitivity = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	Specificity = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Accuracy



bigml

PRIVATE DEPLOYMENTS EVENTS

TAGS

Prediction APIs, Reverse Engineering, Security, Wired

Hype or Reality? Stealing Machine Learning Models via Prediction APIs

by atakancetinsoy on September 30, 2016

Wired magazine just published an article with the interesting title **How to Steal an AI**, where the author explores the topic of reverse engineering Machine Learning algorithms based on a recently published academic paper: **Stealing Machine Learning Models via Prediction APIs**.

Can hacker steal ML model?

Adversarial training

Thank you