

# PRESENTACIÓN

Christian Camilo Urcuqui López

Ing. Sistemas, Magister en Informática y Telecomunicaciones

Big Data Professional

Big Data Scientist

Deep Learning Specialization

Cyber Security Data Scientist, LUMU Technologies

Líder de investigación y desarrollo, laboratorio i2t – U ICESI.

[ulcamilo@gmail.com](mailto:ulcamilo@gmail.com)



Infinity is reachable

## Christian Urcuqui

urcuqui

★ PRO

Edit profile

System engineer, MSc, and Researcher. My research topics are: Data Science, Artificial Intelligence, and Computer Security

Universidad Icesi

ulcamilo@gmail.com

urcuqui.github.com

Overview    **Repositories 47**    Projects 0    Stars 88    Followers 105    Following 41

Find a repository...

Type: All ▾

Language: All ▾

New

### Data-Science

My projects about data science, artificial intelligence and computer security in AI

★ Star



machine-learning    data-science    deep-learning    tensorflow    udacity    busqueda



Jupyter Notebook    ★ 4    18    Updated 11 hours ago

### Ciencia-de-datos-ICESI

Repositorio de la Universidad Icesi para las clases de ciencia de datos de pregrado y posgrado

★ Star



HTML    ★ 4    15    Updated 9 days ago



### WhiteHat

Information about my experiences in ethical hacking

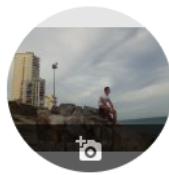
★ Star

security    whitehat    python    hacker    research    researcher    machine-learning



Jupyter Notebook    ★ 41    18    Updated 25 days ago

<https://github.com/urcuqui>



Christian Urcuqui

SEGUIR

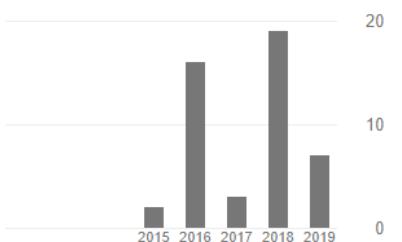
Researcher

Dirección de correo verificada de icesi.edu.co - [Página principal](#)

Cybersecurity Data Science Artificial Intelligence Computer Security Big Data

#### Citado por

	Total	Desde 2014
Citas	47	47
Índice h	4	4
Índice i10	1	1



CITADO POR AÑO

- |                          | TÍTULO  | CITADO POR | AÑO  |
|--------------------------|---|------------|------|
| <input type="checkbox"/> | <a href="#">Machine learning classifiers for android malware analysis</a><br>C Urcuqui, A Navarro<br>Communications and Computing (COLCOM), 2016 IEEE Colombian Conference on, 1-6  | 11 *       | 2016 |
| <input type="checkbox"/> | <a href="#">SafeCandy: System for security, analysis and validation in Android</a><br>S Londoño, C Urcuqui, MF Amaya, J Gómez, AN Cadavid<br>Sistemas y Telemática 13 (35), 89-102  | 9          | 2015 |
| <input type="checkbox"/> | <a href="#">Framework for malware analysis in Android</a><br>C Urcuqui, A Navarro<br>Sistemas & Telemática 14 (37), 45-56   | 8 *        | 2016 |
| <input type="checkbox"/> | <a href="#">Análisis y caracterización de frameworks para detección de aplicaciones maliciosas en Android</a><br>A Navarro Cadavid, S Londoño, CC Urcuqui López, J Gomez<br>Conference: XIV Jornada Internacional de Seguridad Informática ACIS-2014 14 | 7          | 2014 |
| <input type="checkbox"/> | <a href="#">Valoración de la plataforma ASEF como base para detección de malware en aplicaciones Android</a><br>M Fuentes, J Gómez<br>Ingenium 8 (21), 11-23  | 4          | 2014 |
| <input type="checkbox"/> | <a href="#">Machine Learning Classifiers to Detect Malicious Websites</a><br>C Urcuqui, A Navarro, J Osorio, M Garcia<br>CEUR Workshop Proceedings 1950, 14-17  | 3          | 2017 |
| <input type="checkbox"/> | <a href="#">Automated bandwidth measurements using ITU-R SM. 443 and GNU radio devices</a><br>A Navarro, L Vargas, C Urcuqui, J Aristizabal, A Arteaga<br>32nd URSI GASS  | 2          | 2017 |
| <input type="checkbox"/> | <a href="#">Antidefacement-State of art</a><br>CCU López, MG Peña, JLO Quintero, AN Cadavid<br>Sistemas & Telemática 14 (39), 9-27  | 2 *        | 2016 |
| <input type="checkbox"/> | <a href="#">Security control for website defacement</a><br>O Mondragón, AFM Arcos, C Urcuqui, AN Cadavid  | 1          | 2017 |
| <input type="checkbox"/> | <a href="#">LSTM and Convolution Networks exploration for Parkinson's Diagnosis</a><br>JF Reyes, JS Montealegre, YJ Castano, C Urcuqui, A Navarro<br>2019 IEEE Colombian Conference on Communications and Computing (COLCOM), 1-4                       |            | 2019 |

Coautores EDITAR

- |  |   |   |
|--|---|---|
|  | Andres Navarro Cadavid<br>Universidad Icesi             | > |
|  | Javier Diaz-Cely<br>Assistant Professor ICESI Univer... | > |

<https://scholar.google.es/citations?user=q6dRgYIAAAAJ&hl=es>

# CONTENIDO

- Motivación
- Contexto
- Ciencia de datos - ciberseguridad
- Aprendizaje seguro
- Conclusiones



# MOTIVACIÓN

**¿100% de seguridad?**



# MOTIVACIÓN



- Ataques informáticos no registrados o estudiados.
- Malas prácticas en seguridad para el desarrollo de software y hardware.
- Vulnerabilidades en los sistemas.
- Nuevas tecnologías.
- Los insiders.
- Requerimientos no funcionales.



# Hackers Used WhatsApp 0-Day Flaw to Secretly Install Spyware On Phones

May 14, 2019 by Swati Khandelwal



## New Exploit for MikroTik Router WinBox Vulnerability Gives Full Root Access

October 08, 2018 by Swati Khandelwal



FayerWayer. MÓVILES VIDEOJUEGOS CIENCIA INTERNET HARDWARE ENTRETENIMIENTO



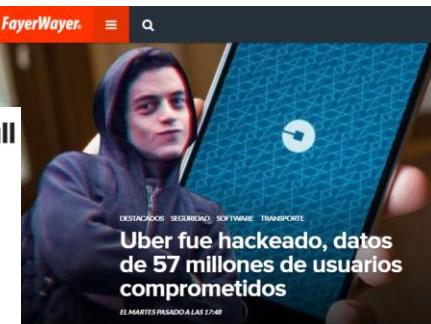
SISTEMAS OPERATIVOS  
Microsoft soluciona el terrible error en la actualización de Windows 10, pero aún no puedes instalarla

HOY 14:37  
El gravísimo error de Windows 10 provocó la pérdida de archivos y Microsoft incluso llegó a pedir que no usaras tu computadora.

39 f t G+ @



por ELIZABETH LEGARRETA



La compañía pagó USD \$100.000 por ocultar el hackeo.

1856 f t G+ @

Se vienen días rudos para Uber, y cualquier persona que haya usado esta plataforma. Más allá de las crisis y renegociaciones



Thursday, May 25, 2017

by Mohit Kumar

G+ 36 Like 1.3K Share 3621 Tweet 1214 in Share 194 Share 5068



EsteemAudit (No Patch)

Windows RDP Hacking Tool

EL PAÍS

ESTADOS UNIDOS

SUSCRIBETE

## Una fuga de datos de Facebook abre una tormenta política mundial

Políticos de EE UU y Reino Unido reclaman que Zuckerberg dé explicaciones tras la revelación de una consultora electoral manipuló información de 50 millones de usuarios de la red social



PABLO DE LLANO | ÁLVARO SÁNCHEZ

Miami / Bruselas • 20 MAR 2018 - 13:26 CDT



Facebook se

el caso Cambi

a que la filtraci

ón de los datos

de los 50 millon

s de usuarios

que se filtraron

en marzo

se ha converti

do en una tem

ática global

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook

que ha llevado

al escándalo

de la red so

cial más grande

de la historia

de Facebook</p

# MOTIVACIÓN



# Complejidad



# MOTIVACIÓN



¿Es posible garantizar el 100% de seguridad?

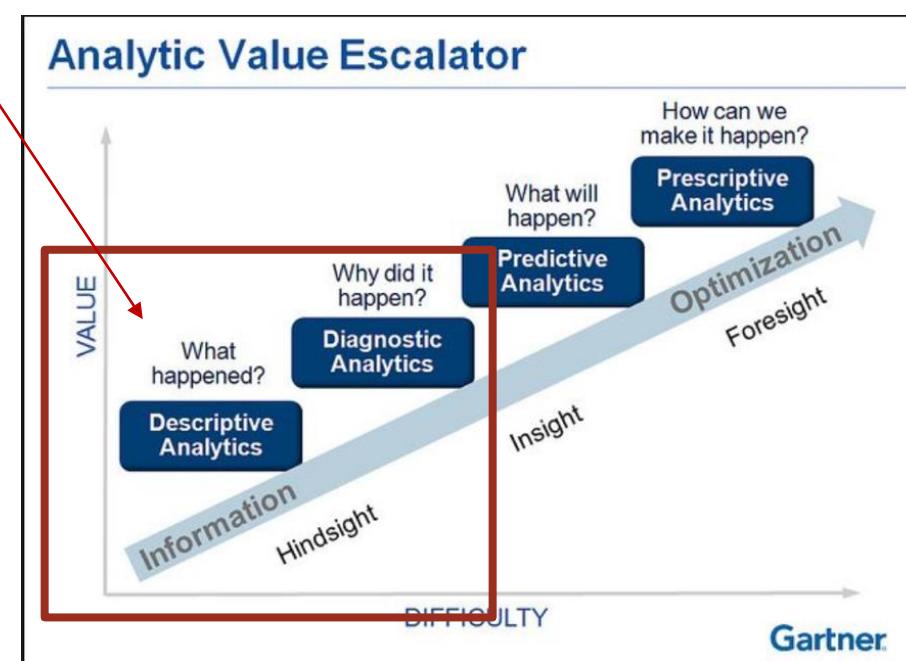
¿Es posible crear una solución determinista que me garantice el 100% de seguridad?

¿Es posible crear un sistema que prediga las acciones de los cibercriminales o de un software malicioso?

# MOTIVACIÓN

Sistemas basados en firmas.

Sistemas basados en anomalías.



Gartner – 2017 Planning Guide For Data Analytics



# MOTIVACIÓN

Las soluciones convencionales de seguridad requieren de un proceso de identificación, es decir, de un esfuerzo humano que implica tiempo y recursos.



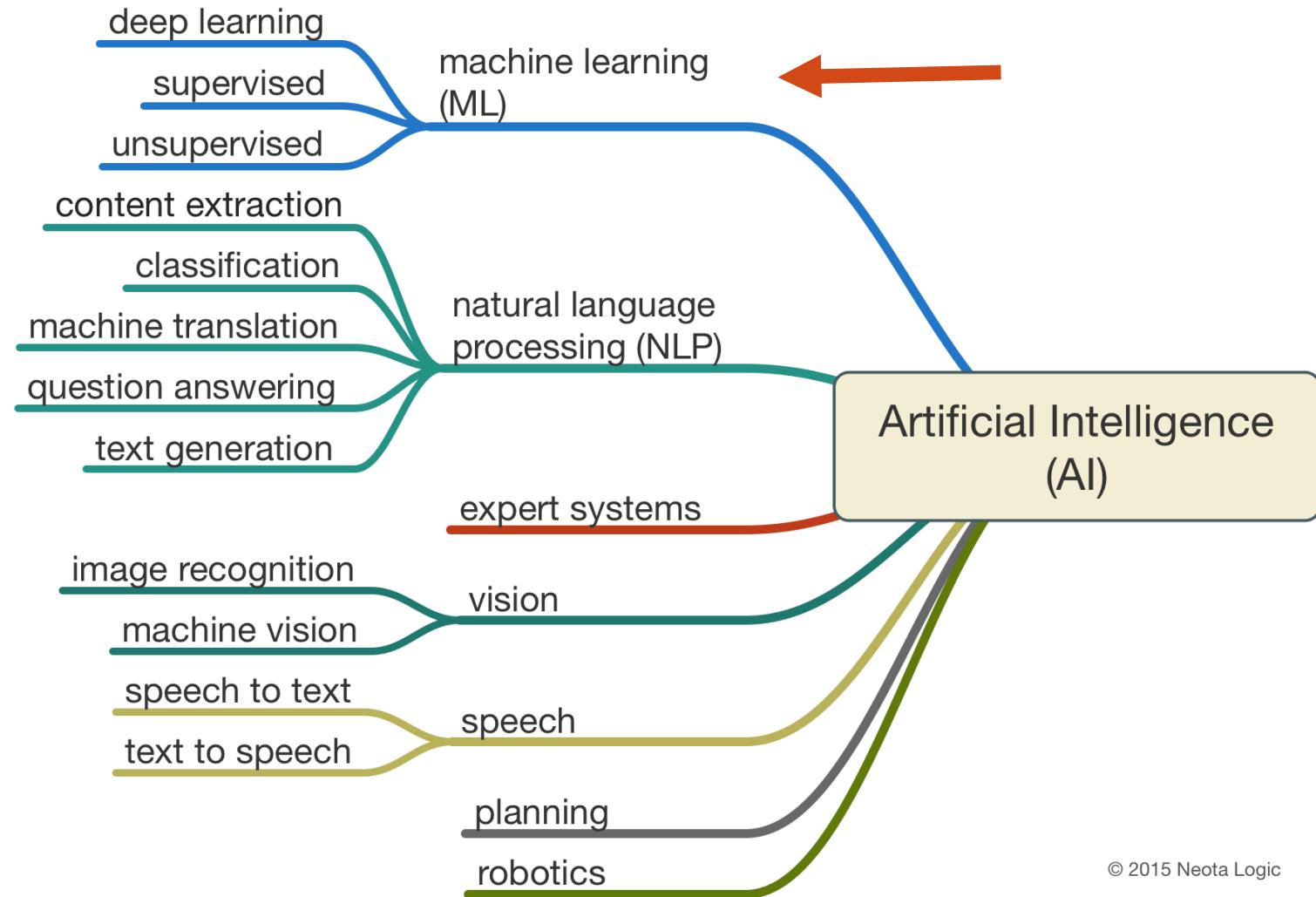
# MOTIVACIÓN

Uno de los caminos más prometedores es la aplicación de algoritmos de aprendizaje de máquina (*machine learning*) que permita ser más eficiente la labor de identificación de nuevas amenazas [1].

[1] Chan, P. K. & Lippmann, R. P. (2006). Machine learning for computer security. *The Journal of Machine Learning Research*, 7, 2669-2672.



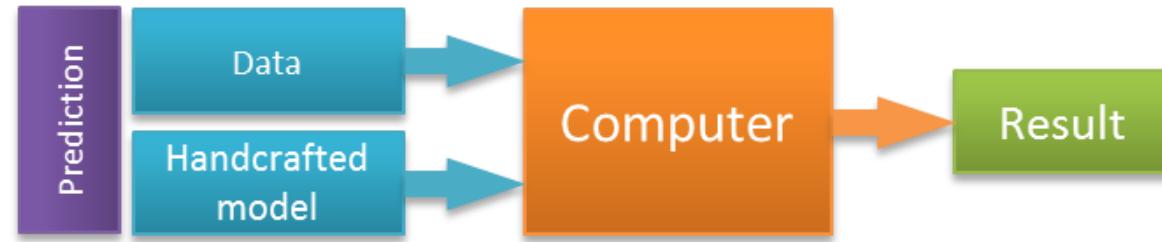
# INTELIGENCIA ARTIFICIAL



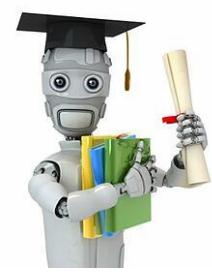
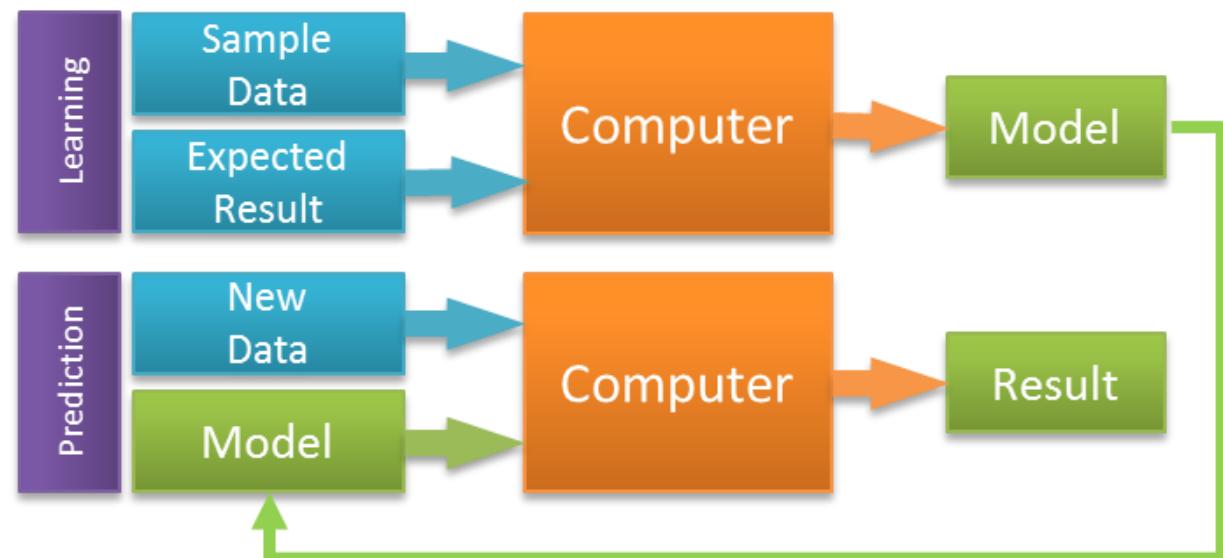
# MACHINE LEARNING

*Machine learning* busca que un sistema tenga la capacidad de aprender en entornos variables, sin que sea programado de forma explícita. Su uso ha venido creciendo debido a los volúmenes de información y a las capacidades computacionales (Big Data).

## Traditional modeling:

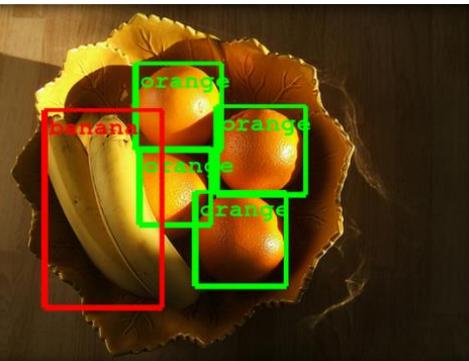
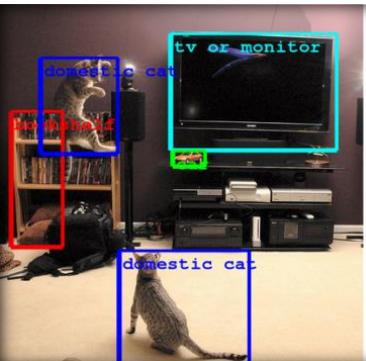
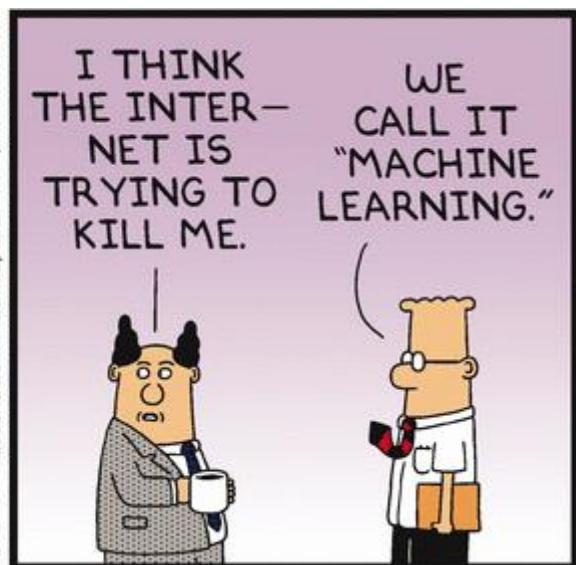


## Machine Learning:



# MACHINE LEARNING

NETFLIX



Dilbert – Machine Learning [2]

[2] Scott, A. <http://dilbert.com/strip/2013-02-02>

# MACHINE LEARNING

Busca que un sistema tenga la capacidad de aprender en entornos variables sin ser programado de forma explícita.

- Aprendizaje supervisado.
- Aprendizaje no supervisado.
- Aprendizaje semisupervisado.
- Aprendizaje por refuerzo.

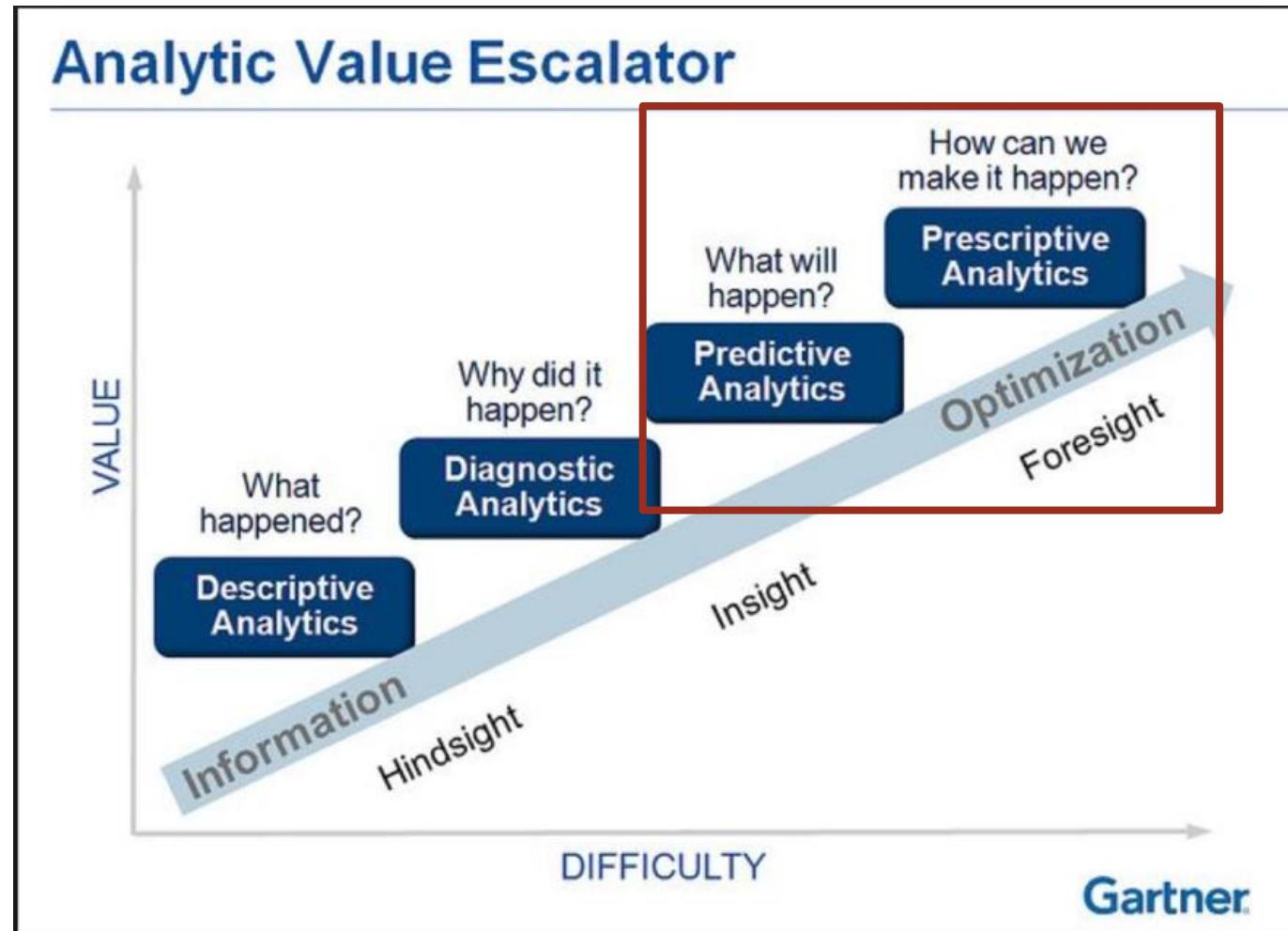


# CIENCIA DE DATOS

La **ciencia de datos** es un área que tiene como objetivo obtener elementos de valor de distintas fuentes de información a través de técnicas y herramientas que incluyen métodos de estadística, minería de datos, *machine learning* y visualización. Es decir, esta área **busca entender y encontrar patrones en los datos con la finalidad de generar modelos que representan al contexto de la información** [4].

[4] Urcuqui, Christian. Ciberseguridad, un enfoque desde la ciencia de datos. (2018). *Editorial Universidad Icesi*

# CIENCIA DE DATOS

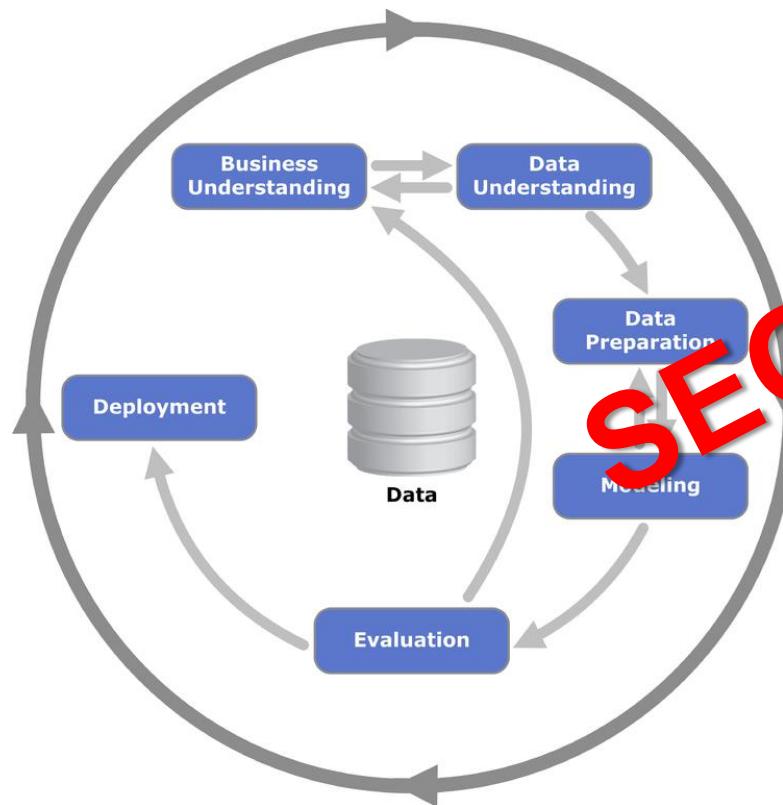


# CIENCIA DE DATOS

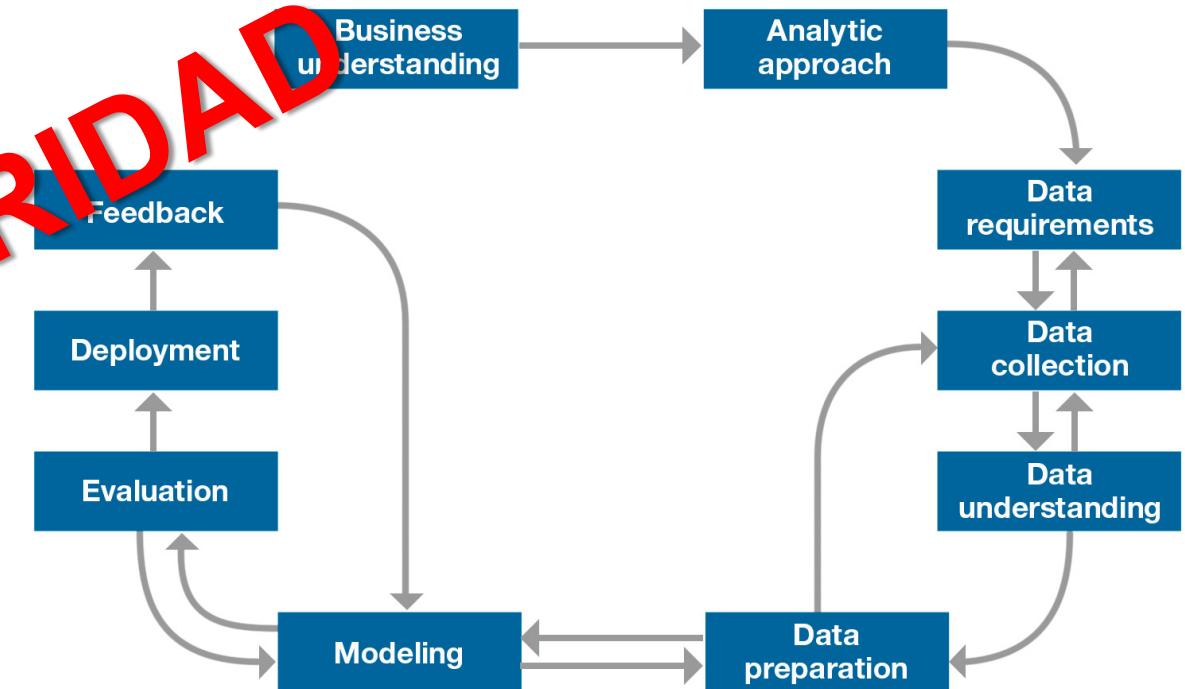


# CIENCIA DE DATOS

CRISP-DM



ASUM-DM



SEGURIDAD

<https://www.ibmdatahub.com/blog/why-we-need-methodology-data-science>

# CIENCIA DE DATOS - CIBERSEGURIDAD



# CIENCIA DE DATOS - CIBERSEGURIDAD



Análisis estático



Análisis dinámico

# CIENCIA DE DATOS - CIBERSEGURIDAD

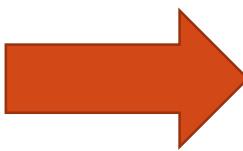


<https://librerianacional.com/producto/ciberseguridad-un-enfoque-desde-la-ciencia-de-datos>

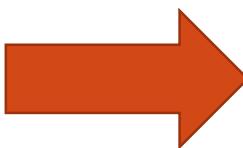
<https://www.icesi.edu.co/editorial/ciberseguridad/>

1. Detección de malware en dispositivos Android
2. Detección de ciberataques web
3. Detección de páginas web maliciosas
4. Controles de seguridad para *Defacement*
5. *Hacking* con hardware
6. Detección de mineros (ilegales) de criptomonedas.
7. Seguridad para la inteligencia artificial (*Adversarial Machine Learning*)

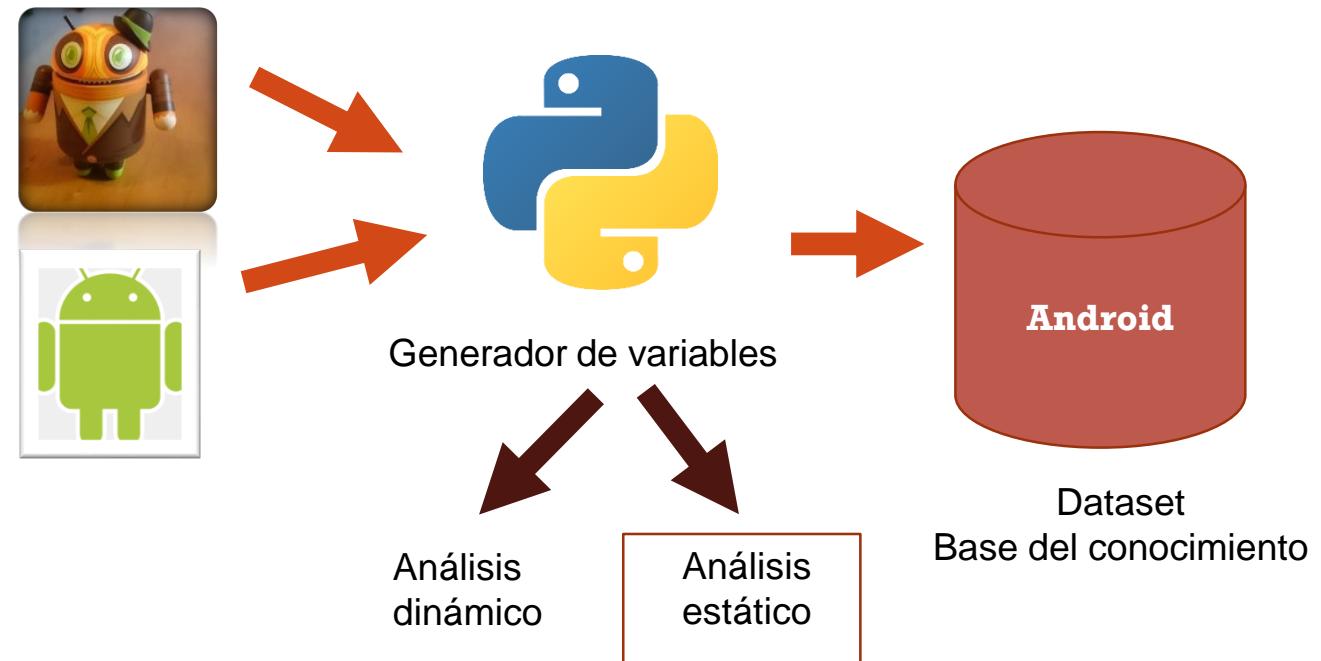
# CIENCIA DE DATOS - CIBERSEGURIDAD



- ¿Cómo clasificar un software benigno y un maligno para dispositivos Android?
- ¿Cómo identificar una página web con contenido maligno?
- ¿Cómo identificar un ciberataque web?
- ¿Cómo vulnerar una inteligencia artificial?
- ¿Cómo detectar una aplicación maliciosa de minería de criptomonedas?
- ¿Cómo detectar deepfake?



# CIENCIA DE DATOS - ANDROID



<https://github.com/urcuqui/whitehat>

# CIENCIA DE DATOS - ANDROID

El generador de variables es un código en Python que recibe una carpeta con archivos APK

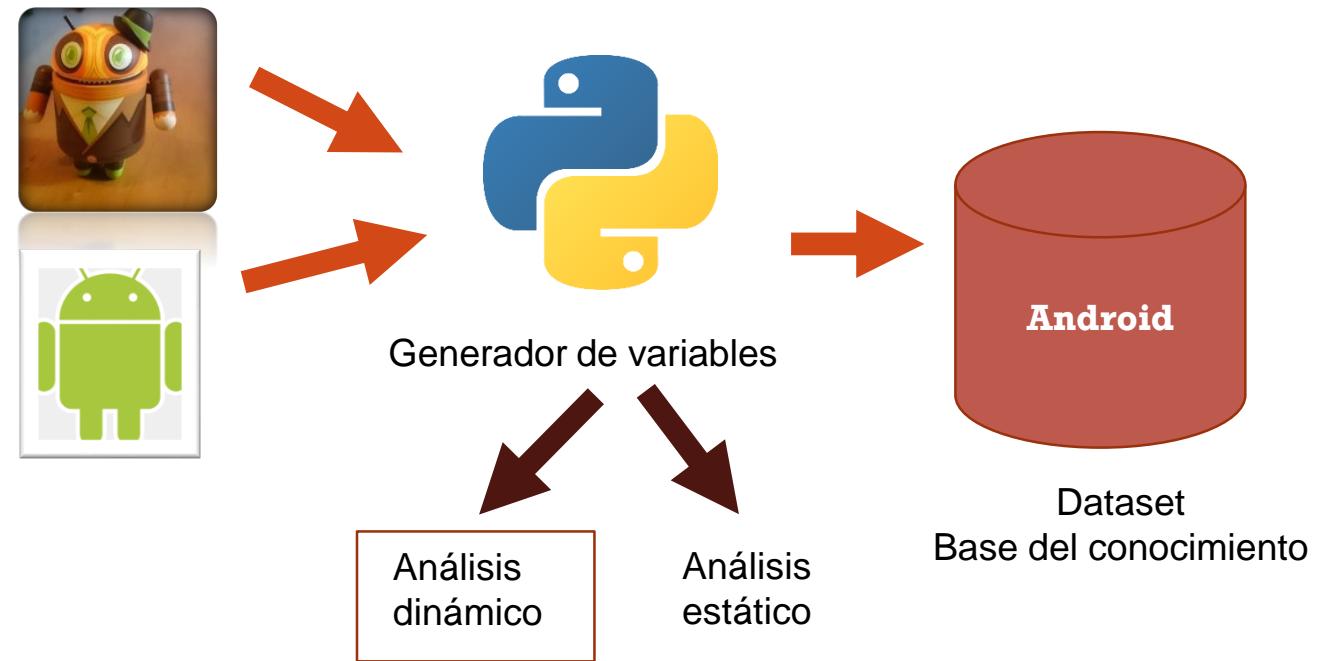
- Utiliza la herramienta ApkTool para obtener los AndroidManifest.xml
- Compara cada archivo AndroidManifest.xml contra una lista de 330 permisos Android generando un dataset

$$R_i = \begin{cases} 1, & \text{Si el analizador detectó un permiso accedido} \\ 0, & \text{En otro caso} \end{cases}$$

$$C_i = \begin{cases} 1, & \text{Si el aplicativo es malicioso} \\ 0, & \text{En otro caso} \end{cases}$$



# CIENCIA DE DATOS - ANDROID



<https://github.com/urcuqui/whitehat>

# CIENCIA DE DATOS - ANDROID

Un sistema de análisis dinámico que se encuentra compuesto por tres módulos:

- Un script que permite ejecutar aplicaciones Android en entornos virtuales (AVD) y generar gestos a través de la herramienta Monkey.
- Un script que captura el tráfico de red de la aplicación mientras es ejecutada en el AVD
- Un script que genera un dataset compuesto por un conjunto de variables del tráfico de red a partir de cada pcap obtenido de la emulación por APK.



# CIENCIA DE DATOS - ANDROID

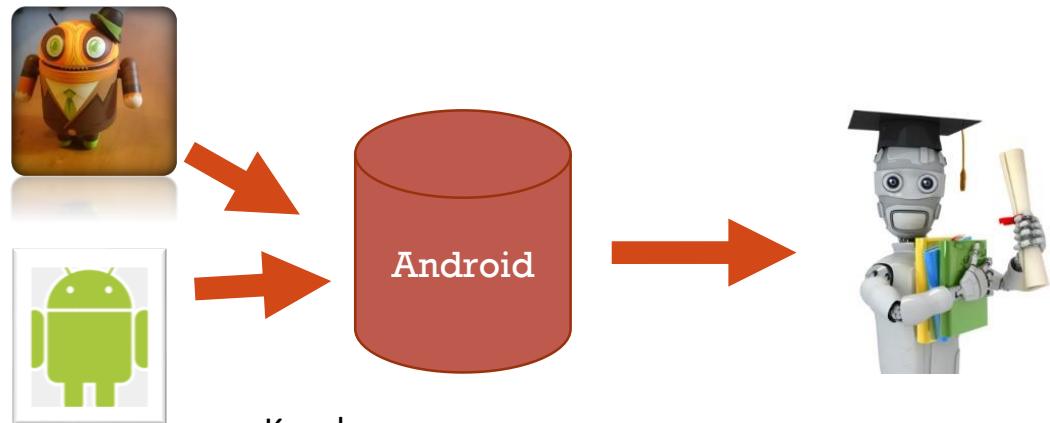
- (R1): *Paquetes TCP*, cuenta el número de paquetes TCP enviados y recibidos durante la comunicación.
- (R2): *Paquetes distintos TCP*, es el número total de paquetes que tienen puertos distintos a los expuestos en TCP.
- (R3): *IP externas*, representa al número direcciones IP externas a las cuales la aplicación se comunicó.
- (R4): *Volumen de bytes*, es el número de bytes que se envía desde la aplicación hacia los sitios externos.
- (R5) *Paquetes UDP*, número total de paquetes UDP transferidos en la comunicación.
- (R6) *Paquetes de la aplicación fuente*, es el número de paquetes enviados desde la aplicación hacia un servidor remoto.
- (R7) *Paquetes de la aplicación remota*, número de paquetes recibidos desde fuentes externas a la aplicación.
- (R8) *Bytes de la aplicación origen*, Este es el volumen (en Bytes) de la comunicación entre la aplicación y el servidor.
- (R9) *Bytes de la aplicación remota*, este es el volumen (en Bytes) de los datos desde el servidor hasta el emulador.
- (R10) *Consultas DNS*, número de consultas DNS.

**7845 registros, 4704 benignos y 3141 maliciosos**

López, C. C. U., Villarreal, J. S. D., Belalcazar, A. F. P., Cadavid, A. N., & Cely, J. G. D. (2018, May). Features to Detect Android Malware. In 2018 IEEE Colombian Conference on Communications and Computing (COLCOM) (pp. 1-6). IEEE.



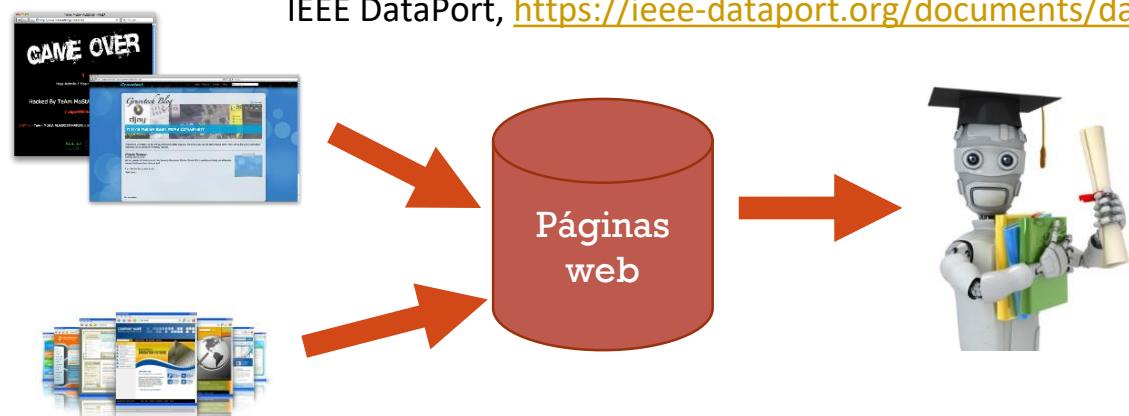
# CIENCIA DE DATOS - DATASETS



Kaggle

- <https://www.kaggle.com/xwolf12/datasetandroidpermissions>
- <https://www.kaggle.com/xwolf12/network-traffic-android-malware>

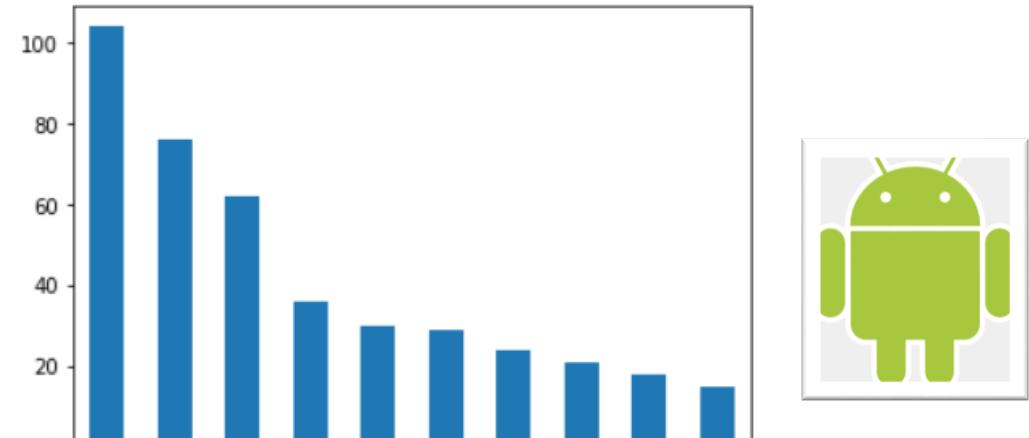
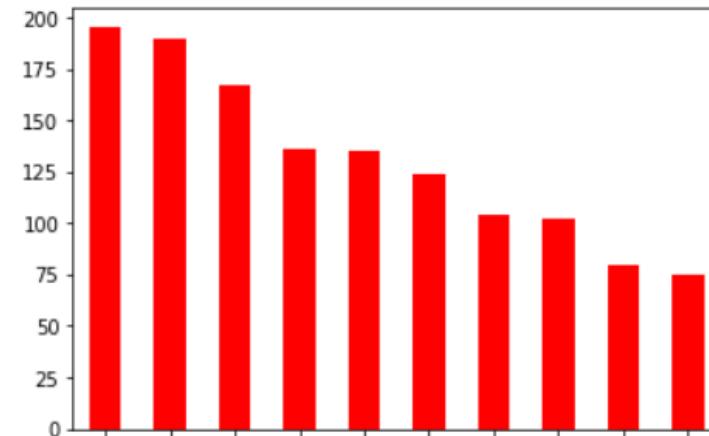
IEEE DataPort, <https://ieee-dataport.org/documents/dataset-malwarebenignn-permissions-android>



Kaggle, <https://www.kaggle.com/xwolf12/malicious-and-benign-websites>

IEEE DataPort, <https://ieee-dataport.org/documents/malicious-and-benign-websites>

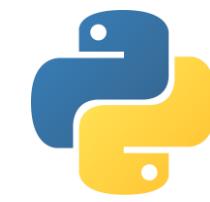
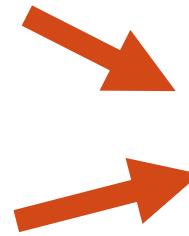
# CIENCIA DE DATOS – ANDROID /199/398



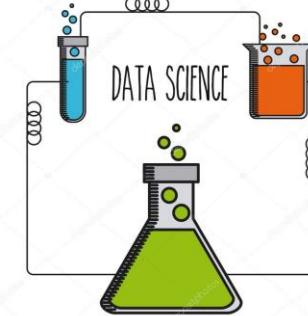
When the flashlight app wants  
access to your call history



# CIENCIA DE DATOS - ANDROID



Generador de variables

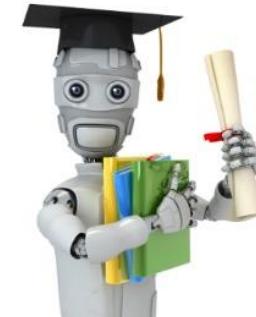


Análisis  
estático

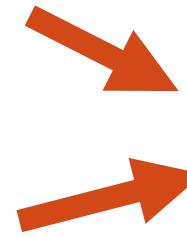
Dataset  
Base del conocimiento



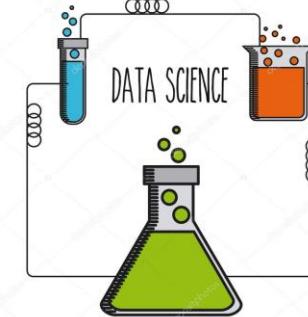
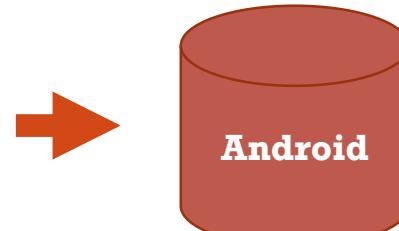
Algoritmo	Metricas de desempeño						
	Precision		Recall		f1 - score		Accuracy
	0	1	0	1	0	1	
NaiveBayes	0,9	0,76	0,79	0,88	0,84	0,82	0,83
Bagging	0,94	0,88	0,88	0,93	0,91	0,9	0,9
Kneighbors = 4	0,94	0,95	0,95	0,94	0,94	0,94	0,94
SVM	0,93	0,88	0,88	0,92	0,9	0,9	0,9
SGD	0,91	0,94	0,94	0,91	0,92	0,93	0,92
DecisionTree	0,93	0,94	0,94	0,93	0,93	0,93	0,93



# CIENCIA DE DATOS - ANDROID



Generador de variables



Análisis  
dinámico

Dataset  
Base del conocimiento



Algoritmo	Metricas de desempeño							
	Precision		Recall		f1 - score		Accuracy	Kappa
	0	1	0	1	0	1		
NaiveBayes	0,81	0,41	0,12	0,96	0,20	0,58	0,44	0,06
Decision Tree	0,90	0,84	0,90	0,85	0,90	0,85	0,87	0,74
Kneighbors = 4	0,89	0,89	0,93	0,83	0,91	0,86	0,89	0,77
SVM	0,62	0,90	1,00	0,06	0,76	0,11	0,62	0,62
RandomForest	0,93	0,90	0,94	0,88	0,93	0,89	0,91	0,82

# APRENDIZAJE SEGURO

Sun Tzu ·



Know the enemy, know yourself;  
your victory will never be  
endangered. Know the ground,  
know the weather; your victory  
will then be total.

AZ QUOTES





# DEEPCODE

## The Democratic Party deepfaked its own chairman to highlight 2020 concerns



By Donie O'Sullivan, CNN Business

Updated 1358 GMT (2158 HKT) August 10, 2019



### TOP STORIES



People are leaving their cars at stations as fuel runs out in Flo



Pompeo gets an unexpected g in home state of Kansas

Recommended by Outbrain

### Paid Content

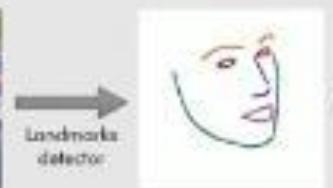


<https://youtu.be/NzI0BkhC4XA>

<https://www.blackhat.com/latestintel/06132019-black-hat-q-and-a-defending-against-deepfakes.html>

## Learning talking heads from few examples

Training frames:



<https://edition.cnn.com/2019/08/09/tech/deepfake-tom-perez-dnc-defcon/index.html?fbclid=IwAR3KXnFp5RhxuD691orvZV3HW-BW5WhYuL59ctC7G85FmdPmeyUMvQePO7Y>

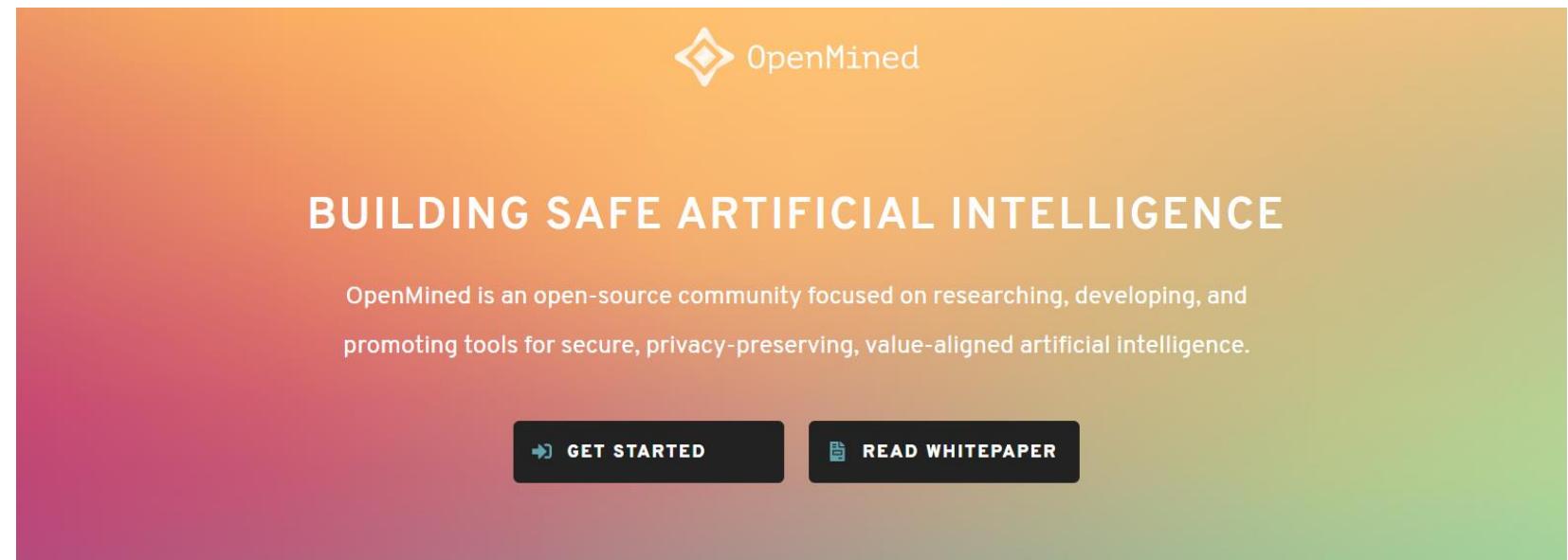
# PRIVACIDAD Y ANONIMIDAD

Es necesario recurrir a mecanismos que garanticen la privacidad y la anonimidad de la información que se utiliza para el entrenamiento de los modelos y del mismo modo no existan fugas de datos sensibles durante el proceso.

Algunas técnicas:

*Differential Privacy*

- The PATE Framework
- Federated Learning



The image shows the homepage of the OpenMined website. The background features a warm, gradient color palette transitioning from orange at the top to red, green, and yellow at the bottom. In the upper left corner, there is a white diamond-shaped logo with a smaller diamond inside it, followed by the text "OpenMined". The center of the page has the text "BUILDING SAFE ARTIFICIAL INTELLIGENCE" in large, bold, white capital letters. Below this, a smaller paragraph reads: "OpenMined is an open-source community focused on researching, developing, and promoting tools for secure, privacy-preserving, value-aligned artificial intelligence." At the bottom, there are two dark call-to-action buttons: one labeled "GET STARTED" with a right-pointing arrow icon, and another labeled "READ WHITEPAPER" with a document icon.





# ADVERSARIAL MACHINE LEARNING

**Attacking Machine Learning with Adversarial Examples**

FEBRUARY 24, 2017

**Adversarial examples** are inputs to machine learning models that an attacker has intentionally designed to cause the model to make a mistake; they're like optical illusions for machines. In this post we'll show how adversarial examples work across different mediums, and will discuss why securing systems against them can be difficult.

At OpenAI, we think adversarial examples are a good aspect of security to work on because they represent a concrete problem in AI safety that can be addressed in the short term, and because fixing them is difficult enough that it requires a serious research effort. (Though we'll need to explore many aspects of machine learning security to achieve our [goal of building safe, widely distributed AI](#).)

<https://blog.openai.com/adversarial-example-research/>

## CLASSIFIERS UNDER ATTACK

Wednesday, February 1, 2017 - 2:00pm-2:30pm

David Evans, University of Virginia



@UdacityDave

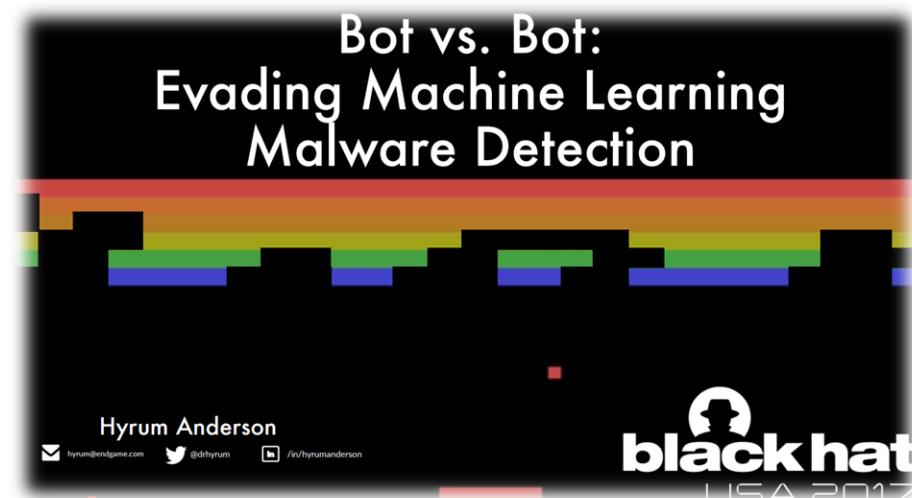
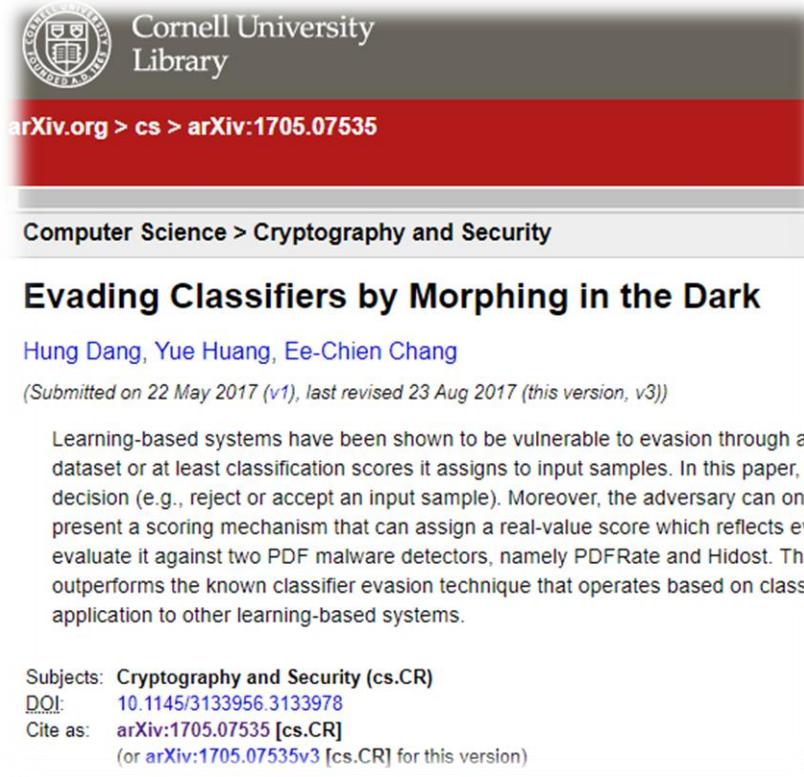
### Abstract:

Machine learning classifiers are widely used in security applications, and often achieve outstanding performance in testing. When deployed, however, classifiers can often be thwarted by motivated adversaries who can construct evasive variants which are misclassified as benign. The main reason for this is that classifiers are trained on samples collected from previous attacks, which often differ from benign samples in superficial and easily-modified ways. Further, many machine learning techniques, including deep neural networks, are inherently fragile. In this talk, I'll highlight the reasons most classifiers can be evaded by motivated adversaries and demonstrate some successful evasion techniques, including ones that can be fully automated. Then, I'll talk about methods that could be used to make classifiers less vulnerable to evasion and to evaluate the robustness of a deployed classifiers in the presence of adversaries.

<https://www.usenix.org/conference/enigma2017/conference-program/presentation/evans>



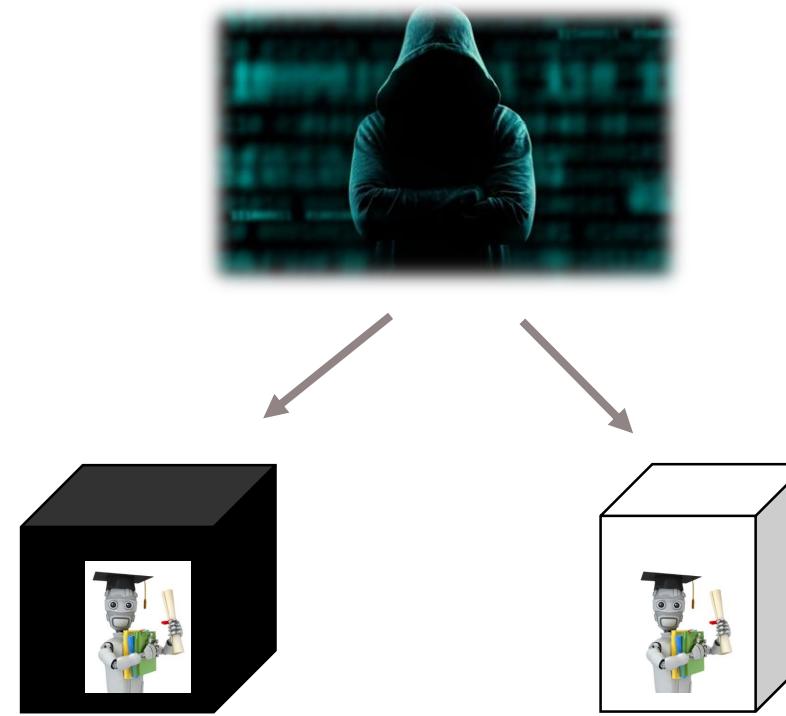
# ADVERSARIAL MACHINE LEARNING



<https://www.blackhat.com/docs/us-17/thursday/us-17-Anderson-Bot-Vs-Bot-Evading-Machine-Learning-Malware-Detection.pdf>

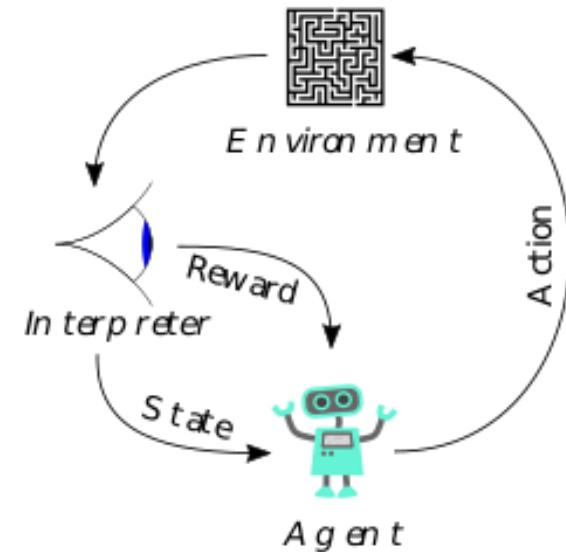
# APRENDIZAJE SEGURO

Se refiere a los algoritmos de aprendizaje que se comportan bien (tienen robustez) en condiciones adversarias



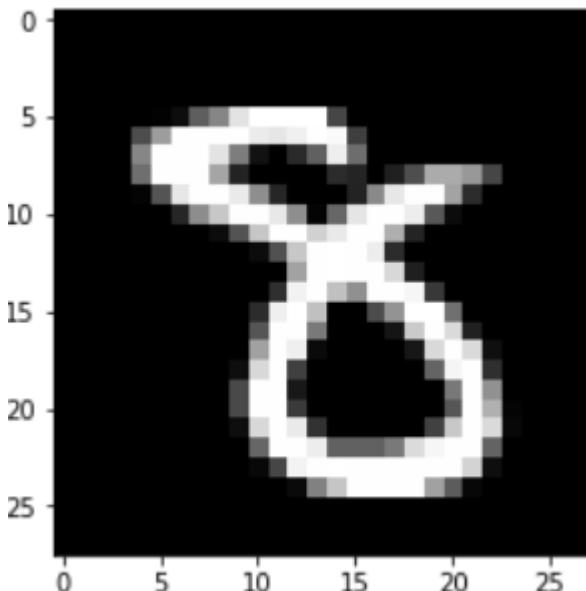
# APRENDIZAJE SEGURO

Un enfoque es un juego de bot versus bot a través del aprendizaje por refuerzo



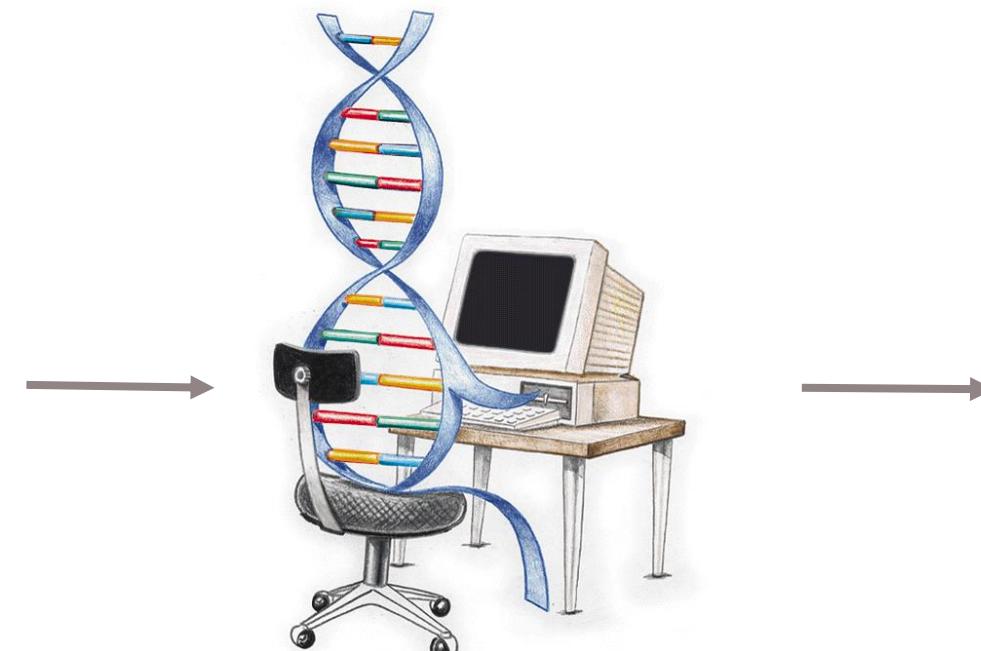
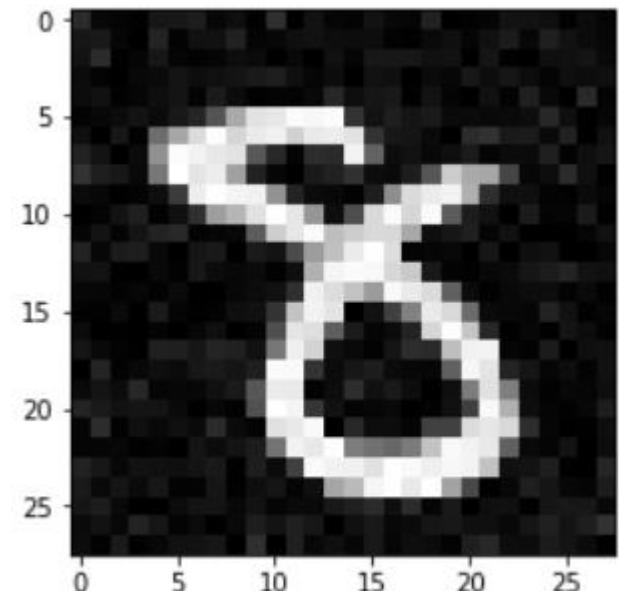
# APRENDIZAJE SEGURO

Antes del ataque

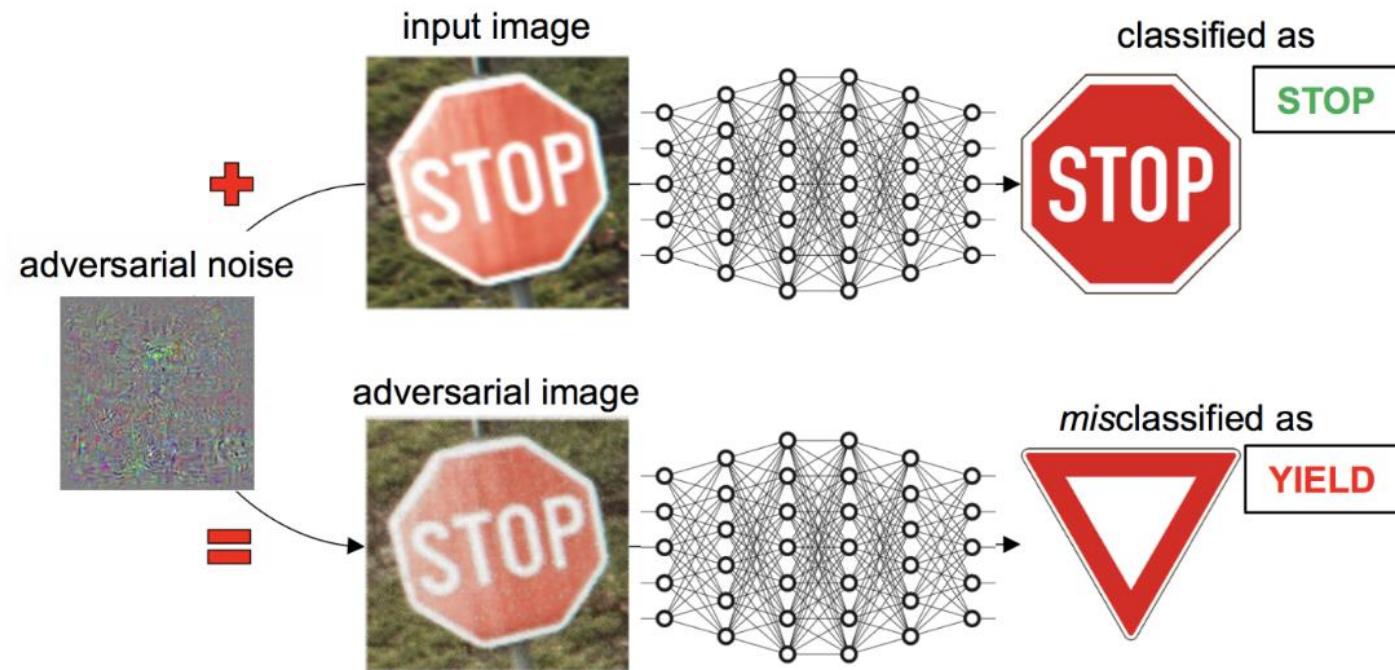


Algoritmo genético

Luego del ataque

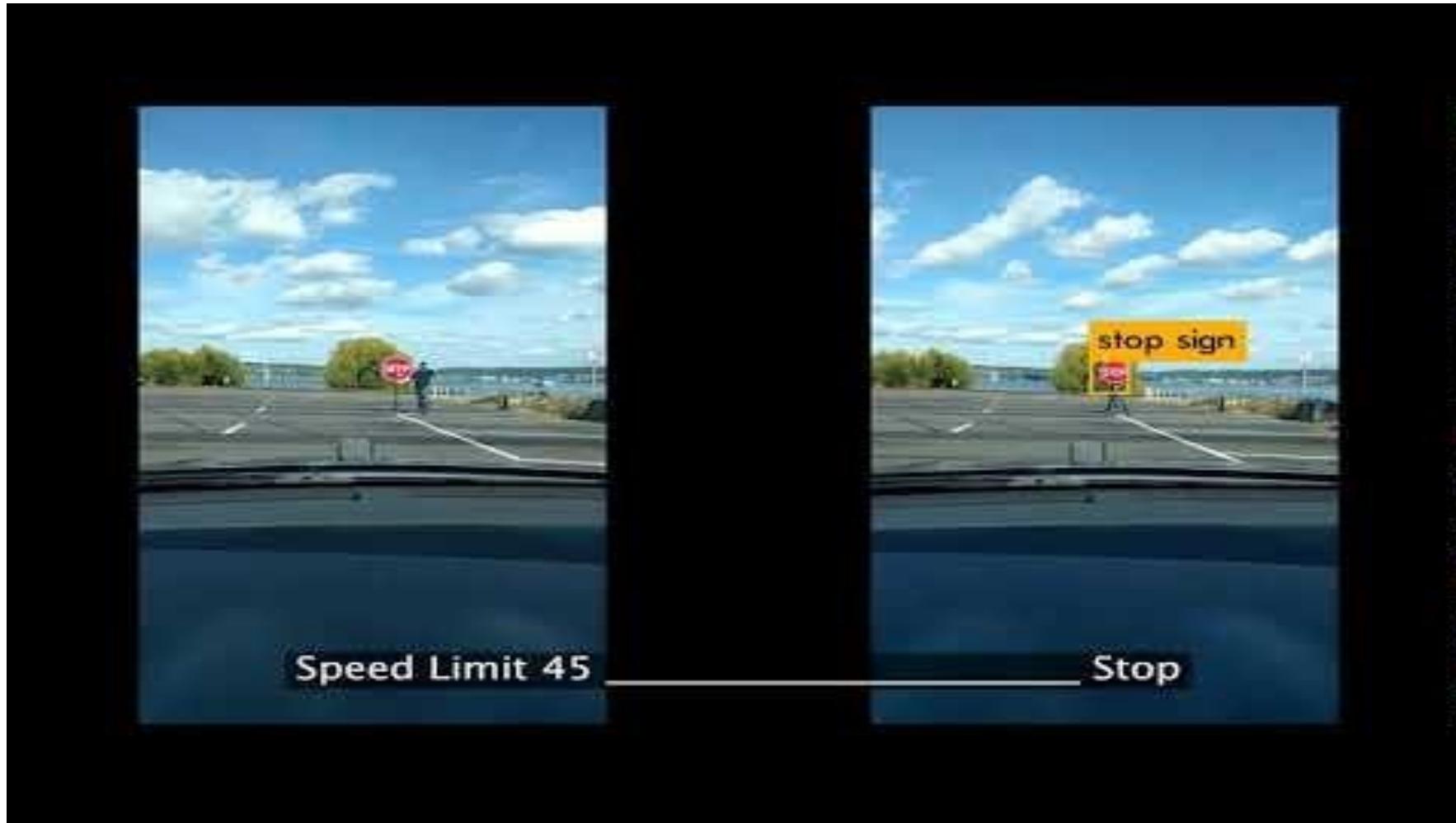


# APRENDIZAJE SEGURO



Battista Biggio, Fabio Roli. Example of adversarial manipulation of an input image. Wild Patterns, Half-day Tutorial on Adversarial Machine Learning



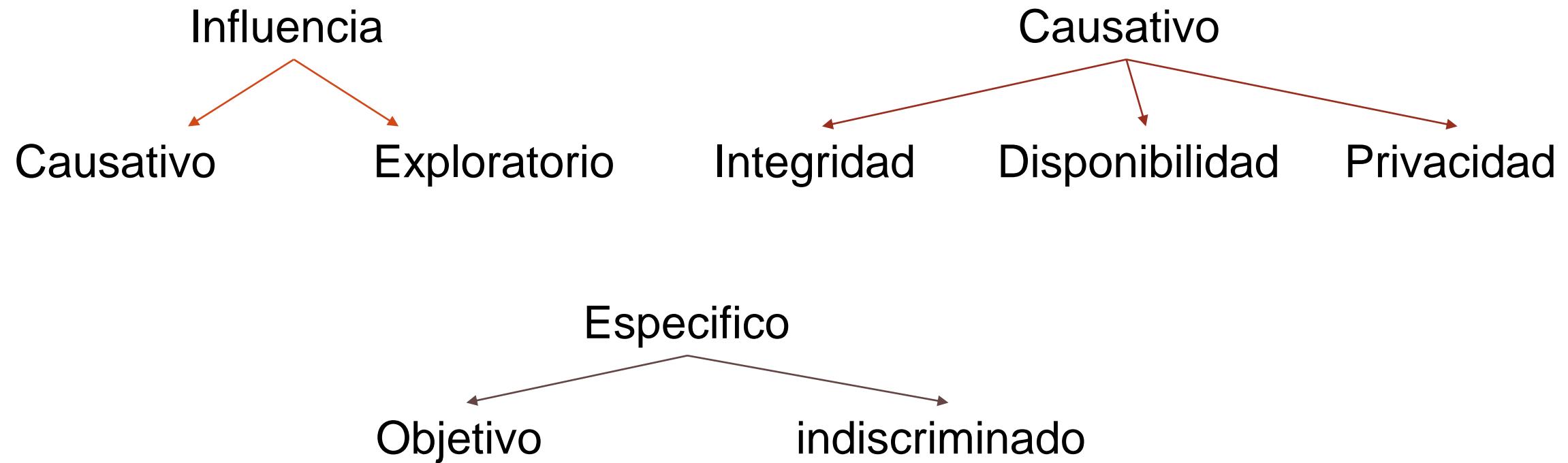


<https://youtu.be/GOjNKQtFs64>



[https://youtu.be/piYnd\\_wYIT8](https://youtu.be/piYnd_wYIT8)

# APRENDIZAJE SEGURO



# APRENDIZAJE SEGURO

## Algunas defensas y contramedidas

### Data sanitization

Defensa RONI (Reject On Negative Impact), una medida que permite estudiar la agregación y la eliminación de datos que tengan un impacto sustancial en la clasificación.

### Robust Learning

Mejorar las capacidades del modelo a partir de ambientes adversarios



# APRENDIZAJE SEGURO



Modelo de defensa basado  
en el tráfico de red

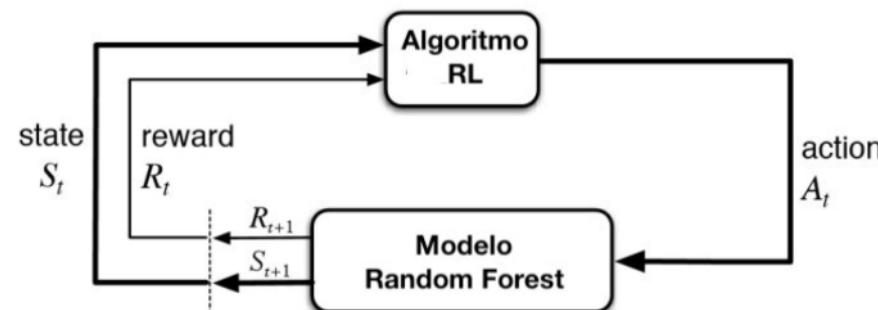


# APRENDIZAJE SEGURO - ATAQUE

## *Enfoque de caja blanca*

Se realizo un experimento sobre una muestra aleatoria de 500 capturas de tráfico de red malicioso.

- **Acciones:**  $A_t = \{R8, R9, R1, R7, R4, R6\}$
- **Reward:**  $R_t = \{+2, 0\}$
- **Perturbación:**  $\min < \alpha_r < \max$



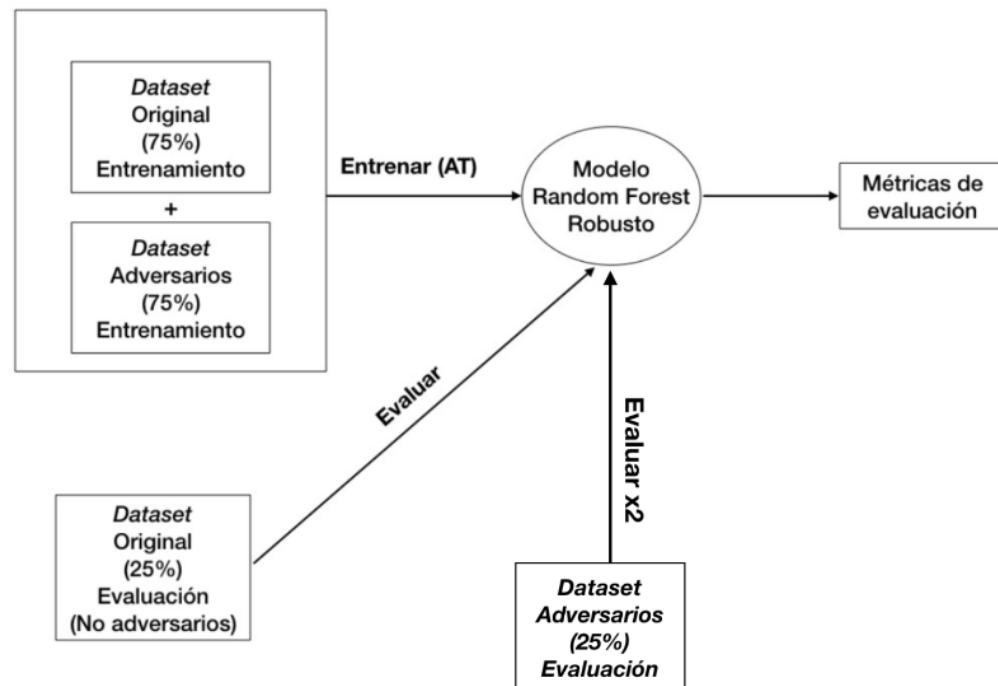
[https://github.com/jsvillatech/SL\\_Android\\_Malware\\_-PDG-](https://github.com/jsvillatech/SL_Android_Malware_-PDG-)

Villarreal, J. S. D., Urcuqui, C., Cadavid, A. N., & Cely, J. (2019). Secure Learning para detección de Android Malware. Tesis de pregrado, ingeniera de sistemas. Universidad Icesi.



# APRENDIZAJE SEGURO

## *Enfoque de caja blanca*



[https://github.com/jsvillatech/SL\\_Android\\_Malware\\_-PDG-/blob/master/Notebooks/Advanced%20Adversarial%20Attack.ipynb](https://github.com/jsvillatech/SL_Android_Malware_-PDG-/blob/master/Notebooks/Advanced%20Adversarial%20Attack.ipynb)

Villarreal, J. S. D., Urcuqui, C., Cadavid, A. N., & Cely, J. (2019). Secure Learning para detección de Android Malware. Tesis de pregrado, ingeniera de sistemas. Universidad Icesi.



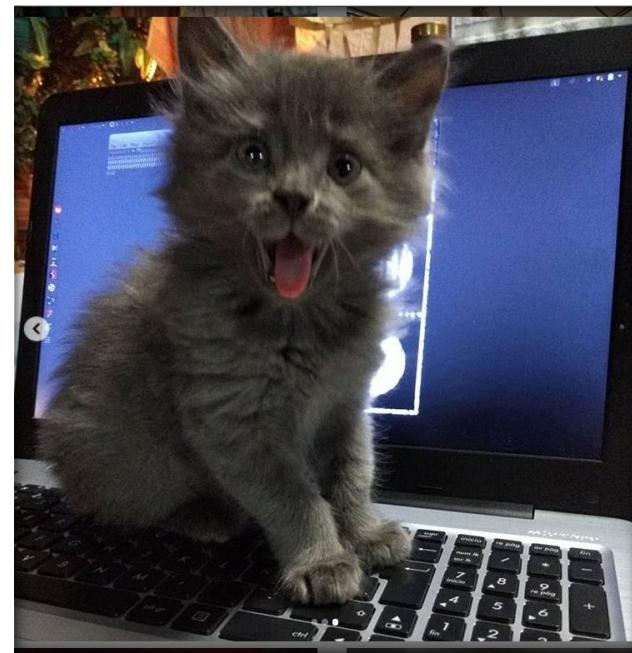
*Resultado en el desempeño individual de los seis modelos*

Algoritmo	Métricas de desempeño							
	Precision		Recall		F1-score		Accuracy	Kappa
	benign	malicious	benign	malicious	benign	malicious	-	-
1. Naive Bayes	0.81	0.41	0.12	0.96	0.20	0.58	0.44	0.06
2. Random Forest	0.93	0.90	0.94	0.88	0.93	0.89	0.91	0.82
3. KNN: K=4	0.89	0.89	0.93	0.83	0.91	0.86	0.89	0.77



Resultado en el desempeño de *Random Forest* luego de realizar *Adversarial training*

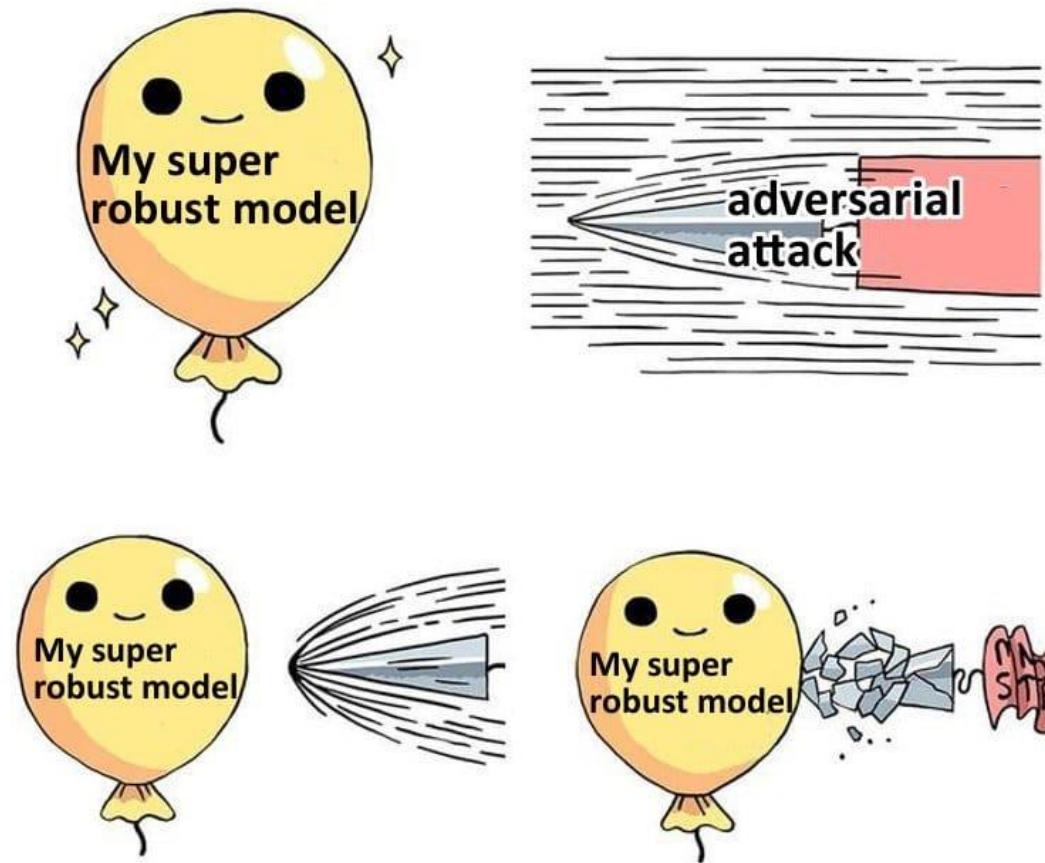
Algoritmo	Métricas de desempeño							
	Precision		Recall		F1-score		Accuracy	Kappa
	benign	malicious	benign	malicious	benign	malicious	-	-
Random Forest Robusto	0.93	0.91	0.94	0.88	0.93	0.89	0.91	0.82



[https://github.com/jsvillatech/SL\\_Android\\_Malware\\_-PDG-/blob/master/Notebooks/Adeversarial%20Training.ipynb](https://github.com/jsvillatech/SL_Android_Malware_-PDG-/blob/master/Notebooks/Adeversarial%20Training.ipynb)

Villarreal, J. S. D., Urcuqui, C., Cadavid, A. N., & Cely, J. (2019). Secure Learning para detección de Android Malware. Tesis de pregrado, ingeniera de sistemas. Universidad Icesi.

# APRENDIZAJE SEGURO





# Muchas gracias

Christian Camilo Urcuqui López  
[ulcamilo@gmail.com](mailto:ulcamilo@gmail.com)

