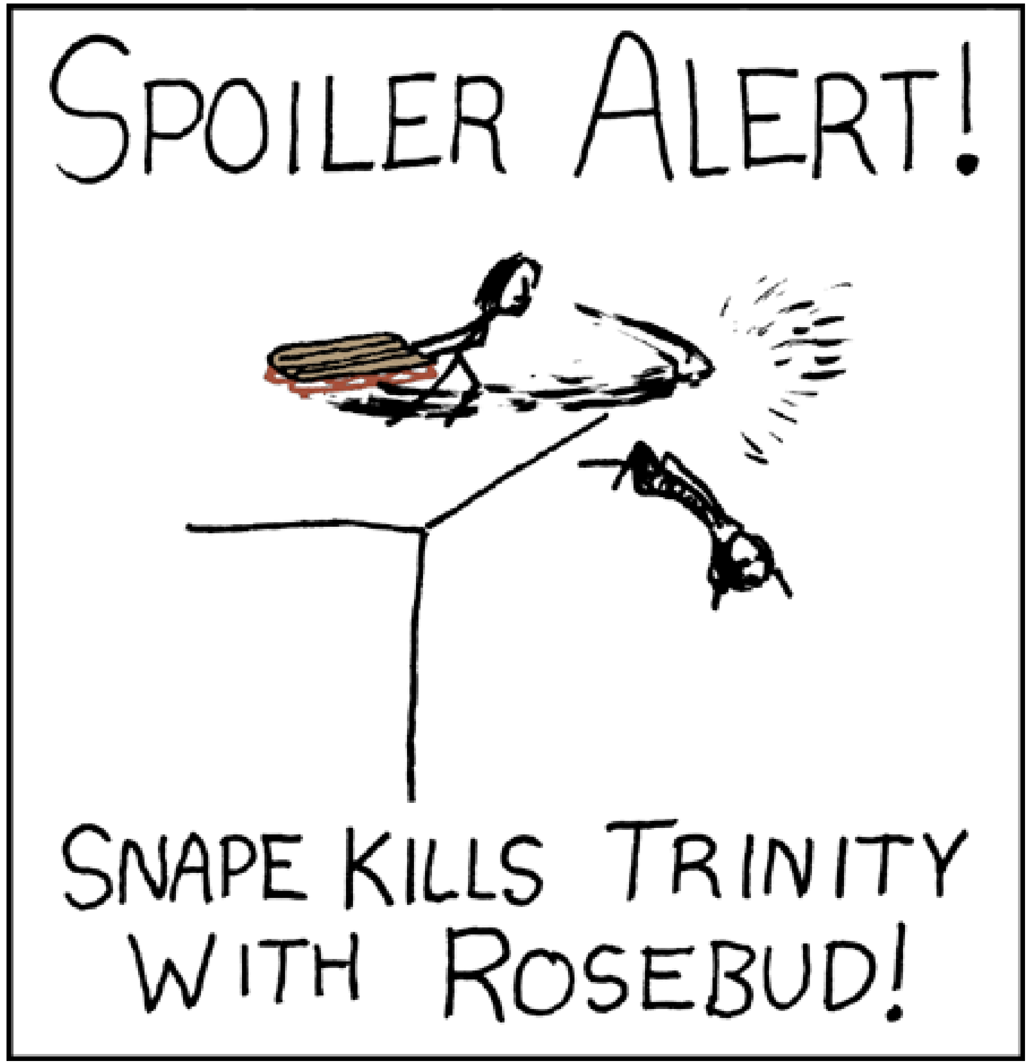
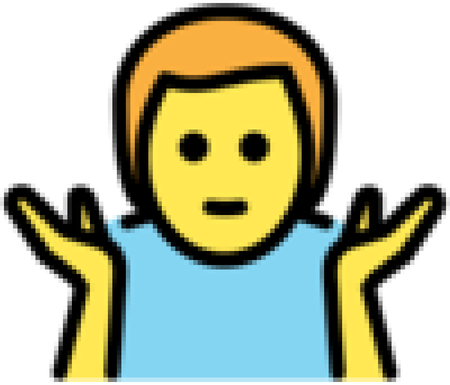
¿Podemos predecir el síndrome visual informático mediante técnicas de machine learning?

Ana Tauste Francés
Rubén Crespo Cano


SPOILER ALERT!
SNAPE KILLS TRINITY
WITH ROSEBUD!

¿Podemos predecir el SVI?

- No lo sabemos (todavía).



El inicio de esta historia...

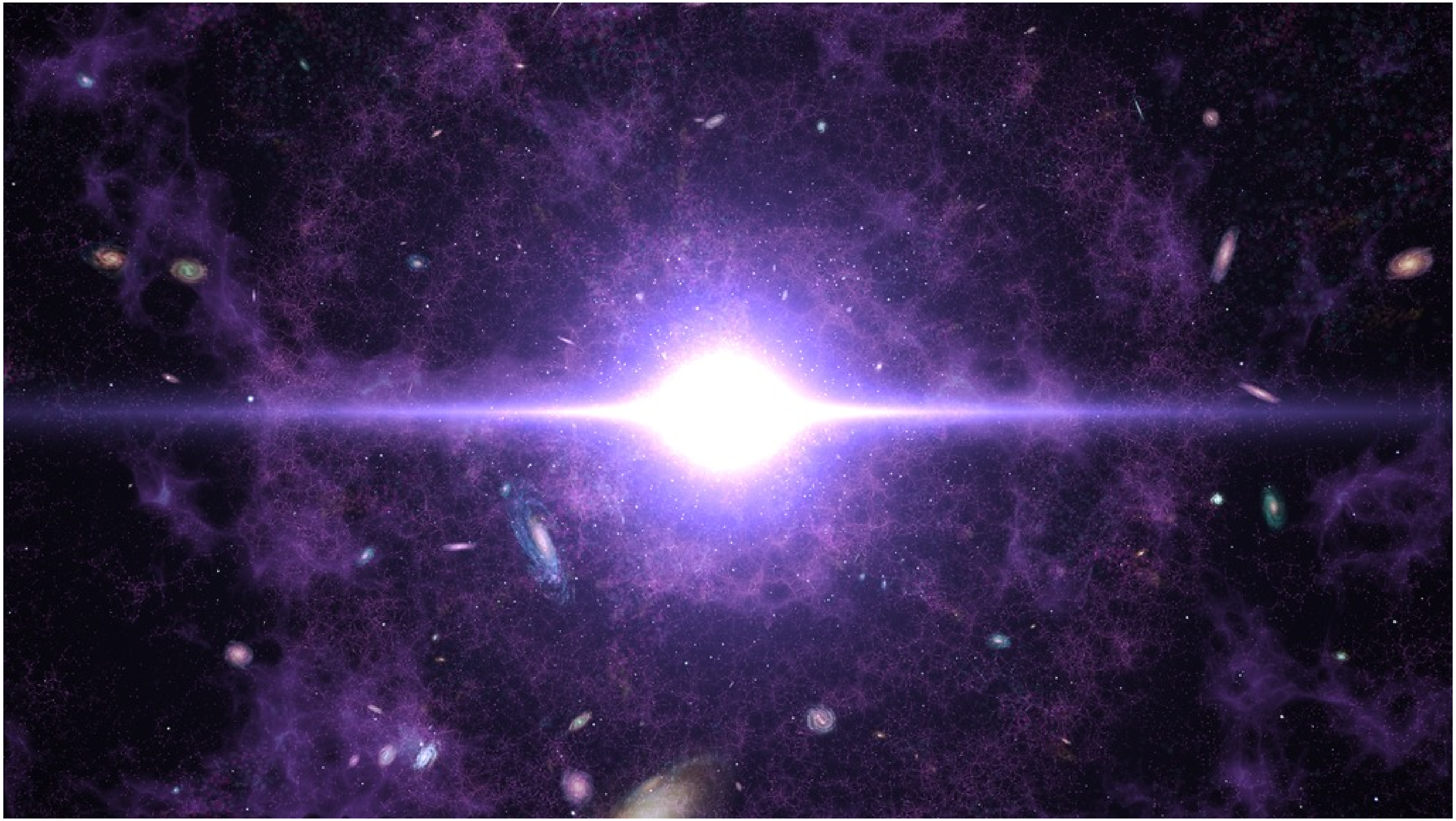


image: NASA's Goddard Space Flight Center/CI Lab

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

A blood-based signature of cerebrospinal fluid Aβ_{1-42} status

Benjamin Goudey[1,2,3], Bowen J. Fung[1,4], Christine Schieber[1], Alzheimer's Disease Metabolomics Consortium*, Alzheimer's Disease Neuroimaging Initiative** & Noel G. Faux[1,5]

Received: 17 April 2018
Accepted: 3 December 2018
Published online: 11 March 2019

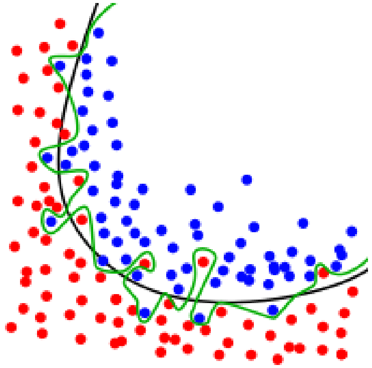
It is increasingly recognized that Alzheimer's disease (AD) exists before dementia is present and that shifts in amyloid beta occur long before clinical symptoms can be detected. Early detection of these molecular changes is a key aspect for the success of interventions aimed at slowing down rates of cognitive decline. Recent evidence indicates that of the two established methods for measuring amyloid, a decrease in cerebrospinal fluid (CSF) amyloid β_{1-42} (Aβ_{1-42}) may be an earlier indicator of Alzheimer's disease risk than measures of amyloid obtained from Positron Emission Tomography (PET). However, CSF collection is highly invasive and expensive. In contrast, blood collection is routinely performed, minimally invasive and cheap. In this work, we develop a blood-based signature that can provide a cheap and minimally invasive estimation of an individual's CSF amyloid status using a machine learning approach. We show that a Random Forest model derived from plasma analytes can accurately predict subjects as having abnormal (low) CSF Aβ_{1-42} levels indicative of AD risk (0.84 AUC, 0.78 sensitivity, and 0.73 specificity). Refinement of the modeling indicates that only *APOE*ε*4* carrier status and four plasma analytes (CGA, Aβ_{1-42}, Eotaxin 3, APOE) are required to achieve a high level of accuracy. Furthermore, we show across an independent validation cohort that individuals with predicted abnormal CSF Aβ_{1-42} levels transitioned to an AD diagnosis over 120 months significantly faster than those with predicted normal CSF Aβ_{1-42} levels and that the resulting model also validates reasonably across PET Aβ_{1-42} status (0.78 AUC). This is the first study to show that a machine learning approach, using plasma protein levels, age and *APOE*ε*4* carrier status, is able to predict CSF Aβ_{1-42} status, the earliest risk indicator for AD, with high accuracy.

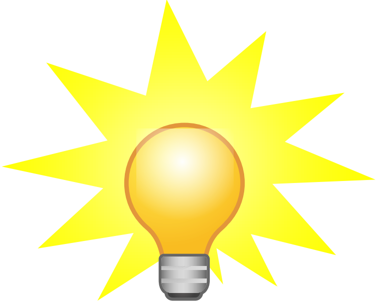
https://doi.org/10.1038/s41598-018-37149-7

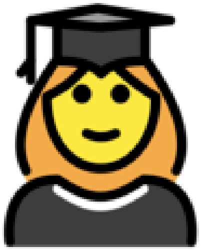
Idea feliz

@rcrespocano

+







@anatauste_

+

Empecemos por el principio...

- ¿Síndrome Visual Informático?
- ¿Machine Learning?

¿Qué es un síndrome?

- *“En medicina, un síndrome [...] es un cuadro clínico o un **conjunto sintomático** que presenta alguna enfermedad con cierto significado y que por sus propias características **posee cierta identidad**; es decir, un **grupo significativo de síntomas y signos** [...], **que concurren en tiempo y forma**, y con variadas causas o etiología [1].”*.

[1] https://es.wikipedia.org/wiki/Síndrome

¿Qué es el Síndrome Visual Informático?

- Conjunto de síntomas oculares, visuales y musculo-esqueléticos que se asocian al uso de dispositivos electrónicos.

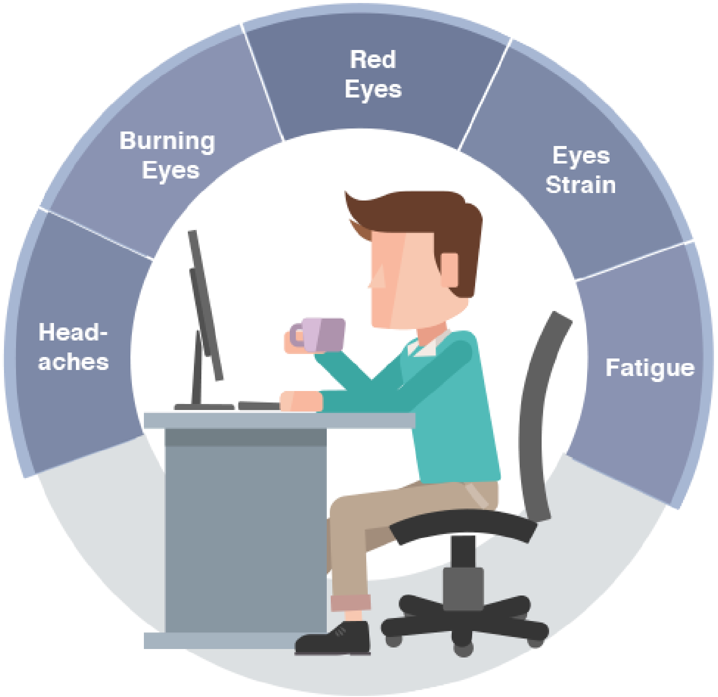


image: ASUS

Síntomas más comunes SVI

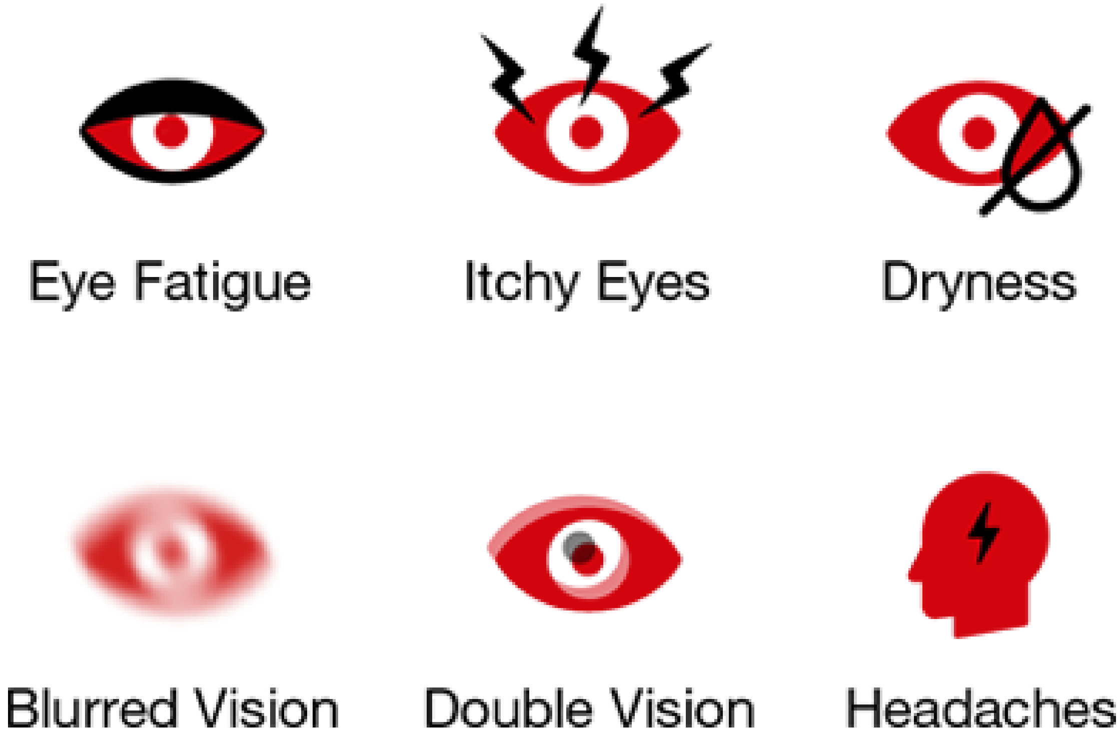


image: https://www.pngkey.com/maxpic/u2q8t4t4t4r5t4q8/

¿Qué factores de riesgo intervienen en el SVI?

- **Factores ambientales**
  - Iluminación, humedad, posición pantalla, etc.

- **Factores de usuario**
  - Sexo, edad, defectos refractivos, uso de lentes de contacto, alteraciones oculo-motoras o de la lágrima, etc.

¿Cómo se mide el SVI?


ELSEVIER

CrossMark

Journal of Clinical Epidemiology 68 (2015) 662–673

Journal of Clinical Epidemiology

A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace

María del Mar Seguí[a,*], Julio Cabrero-García[b], Ana Crespo[a], José Verdú[c], Elena Ronda[d]

[a]Optic Pharmacology and Anatomy Department, Public Health Research Group, University of Alicante, Carretera de San Vicente del Raspeig s/n. 03690, San Vicente del Raspeig, Alicante, Spain
[b]Nursing Department, Faculty of Health Sciences, University of Alicante, Carretera de San Vicente del Raspeig s/n. 03690, San Vicente del Raspeig, Alicante, Spain
[c]Balmis Research Group on Community Health and History of Science, Community Nursing Preventive Medicine Public Health and History of Science Department, University of Alicante, Carretera de San Vicente del Raspeig s/n. 03690, San Vicente del Raspeig, Alicante, Spain
[d]Community Nursing Preventive Medicine Public Health and History of Science Department, Public Health Research Group, Occupational Health Research Centre (CISAL), CIBER Epidemiology and Public Health (CIBERESP), University of Alicante, Carretera de San Vicente del Raspeig s/n. 03690, San Vicente del Raspeig, Alicante, Spain

Accepted 21 January 2015; Published online 28 January 2015

Abstract

Objectives: To design and validate a questionnaire to measure visual symptoms related to exposure to computers in the workplace.
Study Design and Setting: Our computer vision syndrome questionnaire (CVS-Q) was based on a literature review and validated through discussion with experts and performance of a pretest, pilot test, and retest. Content validity was evaluated by occupational health, optometry, and ophthalmology experts. Rasch analysis was used in the psychometric evaluation of the questionnaire. Criterion validity was determined by calculating the sensitivity and specificity, receiver operator characteristic curve, and cutoff point. Test−retest repeatability was tested using the intraclass correlation coefficient (ICC) and concordance by Cohen's kappa (κ).
Results: The CVS-Q was developed with wide consensus among experts and was well accepted by the target group. It assesses the frequency and intensity of 16 symptoms using a single rating scale (symptom severity) that fits the Rasch rating scale model well. The questionnaire has sensitivity and specificity over 70% and achieved good test−retest repeatability both for the scores obtained [ICC = 0.802; 95% confidence interval (CI): 0.673, 0.884] and CVS classification (κ = 0.612; 95% CI: 0.384, 0.839).
Conclusion: The CVS-Q has acceptable psychometric properties, making it a valid and reliable tool to control the visual health of computer workers, and can potentially be used in clinical trials and outcome research. © 2015 Elsevier Inc. All rights reserved.

https://doi.org/10.1016/j.jclinepi.2015.01.015

CVS-Q

- Cuestionario validado.
- 16 síntomas: *visión borrosa, dolor de cabeza, sequedad ocular, lagrimeo, etc.*
- Frecuencia (F): *nunca (0), ocasionalmente (1), a menudo/siempre (2).*
- Intensidad (I): *ninguna (0), moderada (1), intensa (2).*
- Padece SVI si `score ≥ 6`

$$score = \sum_{i=1}^{16} F_i \cdot I_i$$

Síndrome Visual Informático

image: https://flic.kr/p/7Epkxz; CC BY-NC-ND 2.0

Machine Learning

- *“Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.”*
  - Arthur Samuel, 1959

Machine Learning

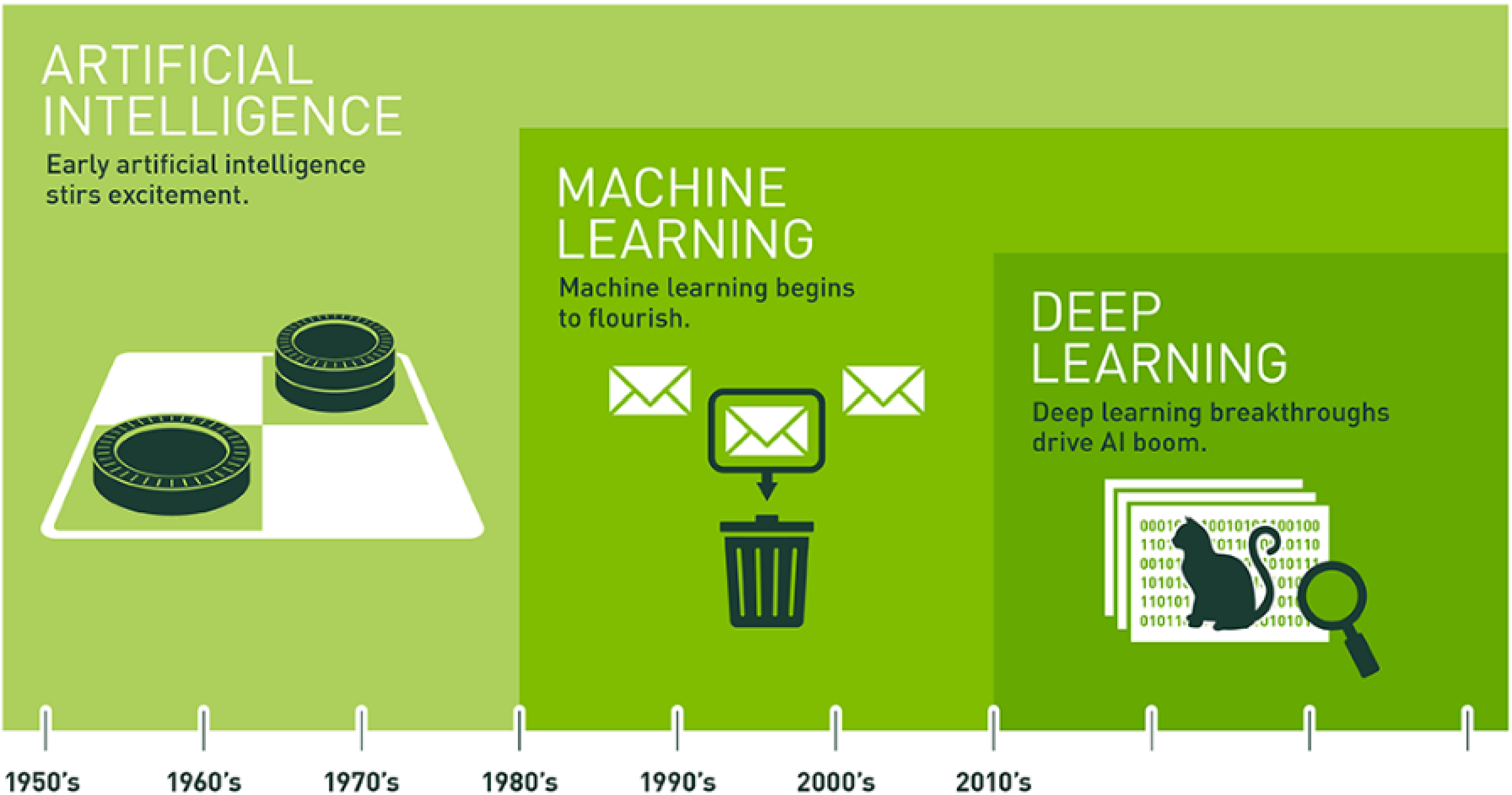
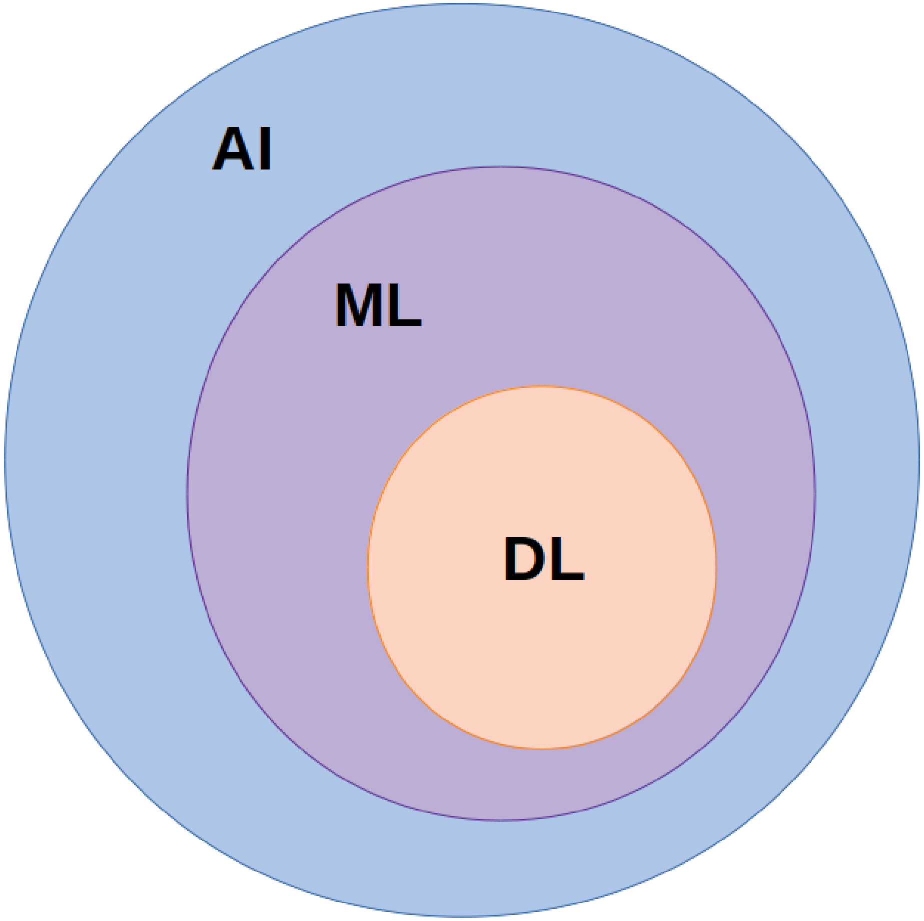
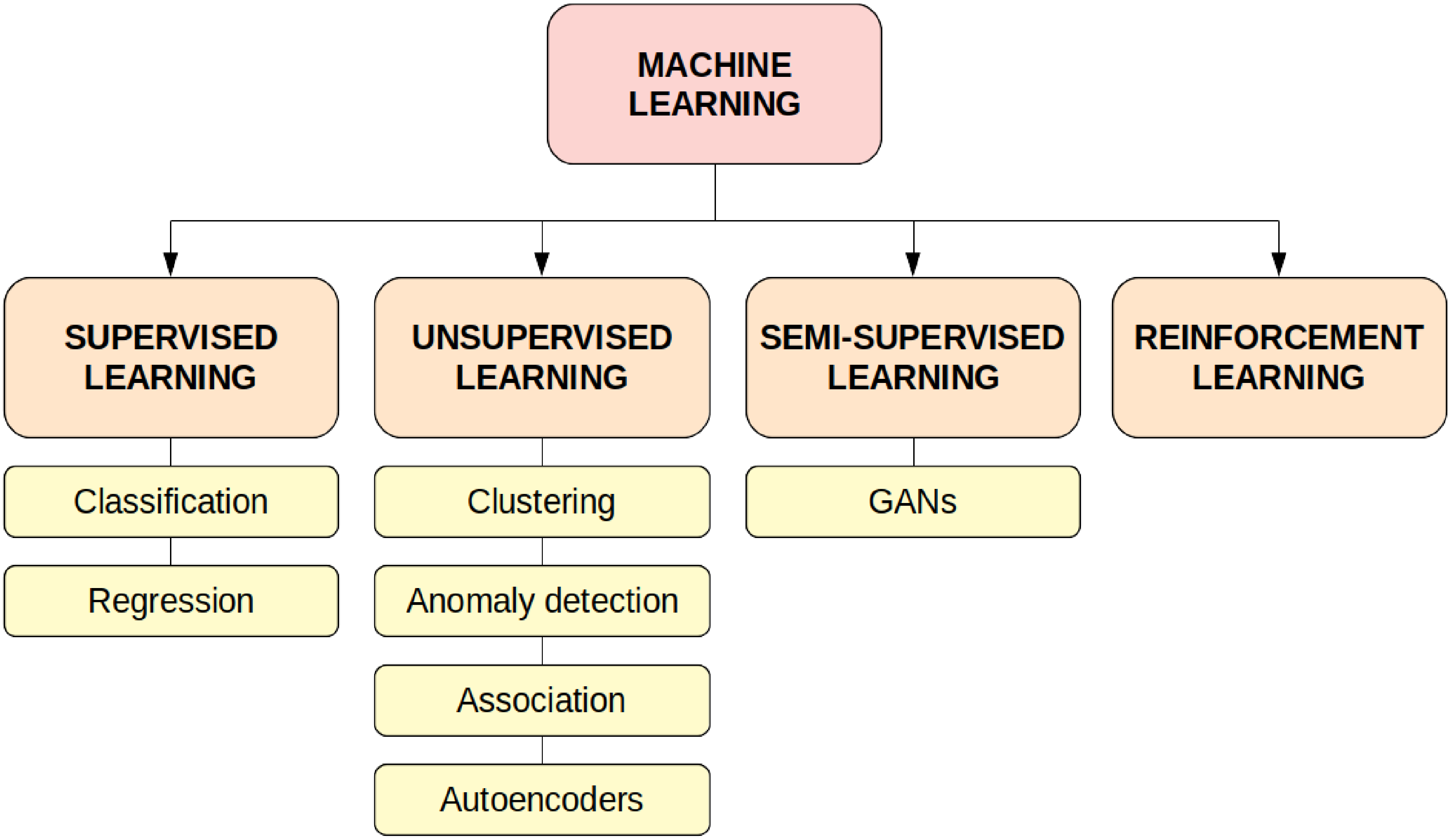


image: NVIDIA; https://developer.nvidia.com/deep-learning

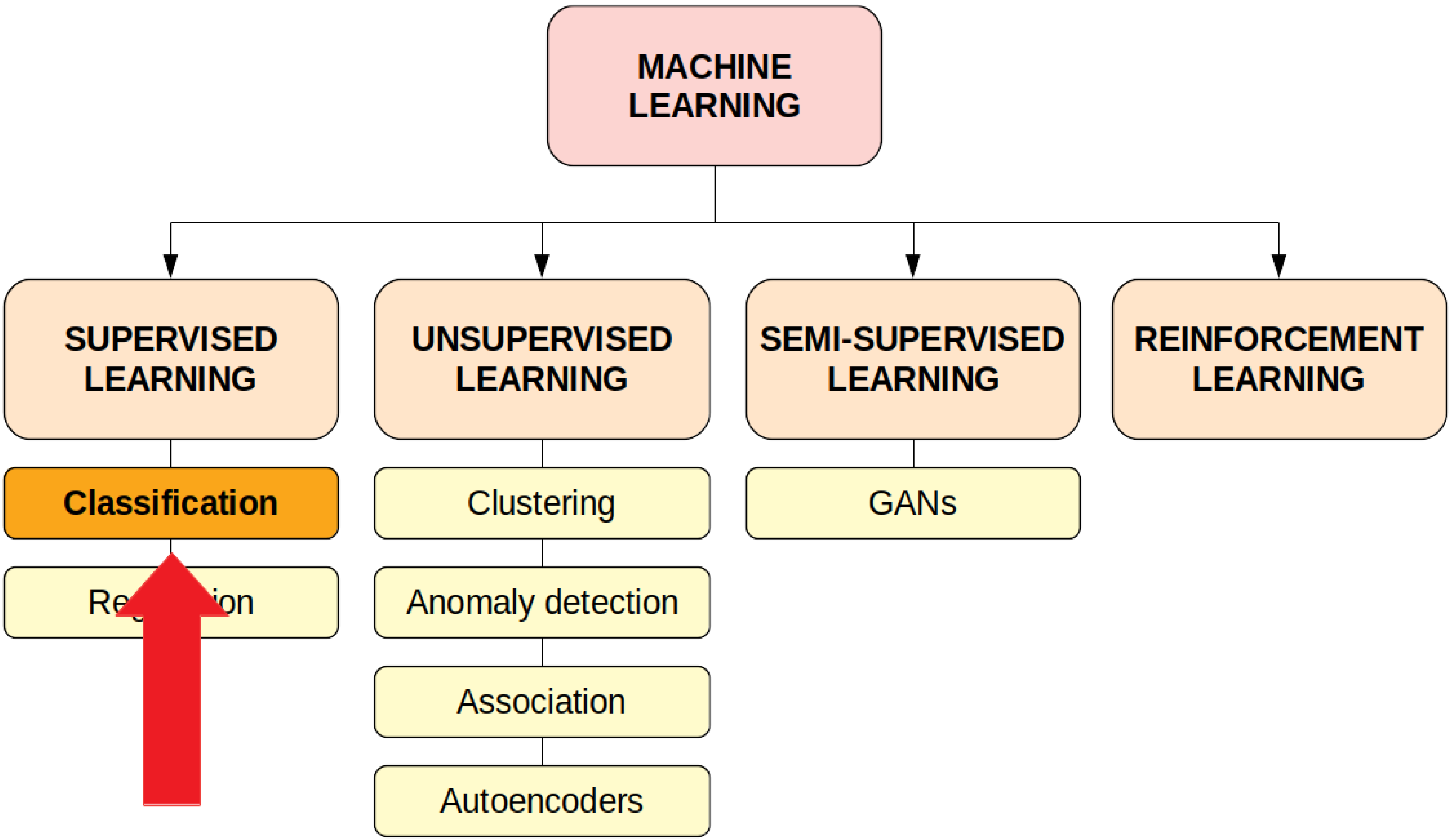
Machine Learning



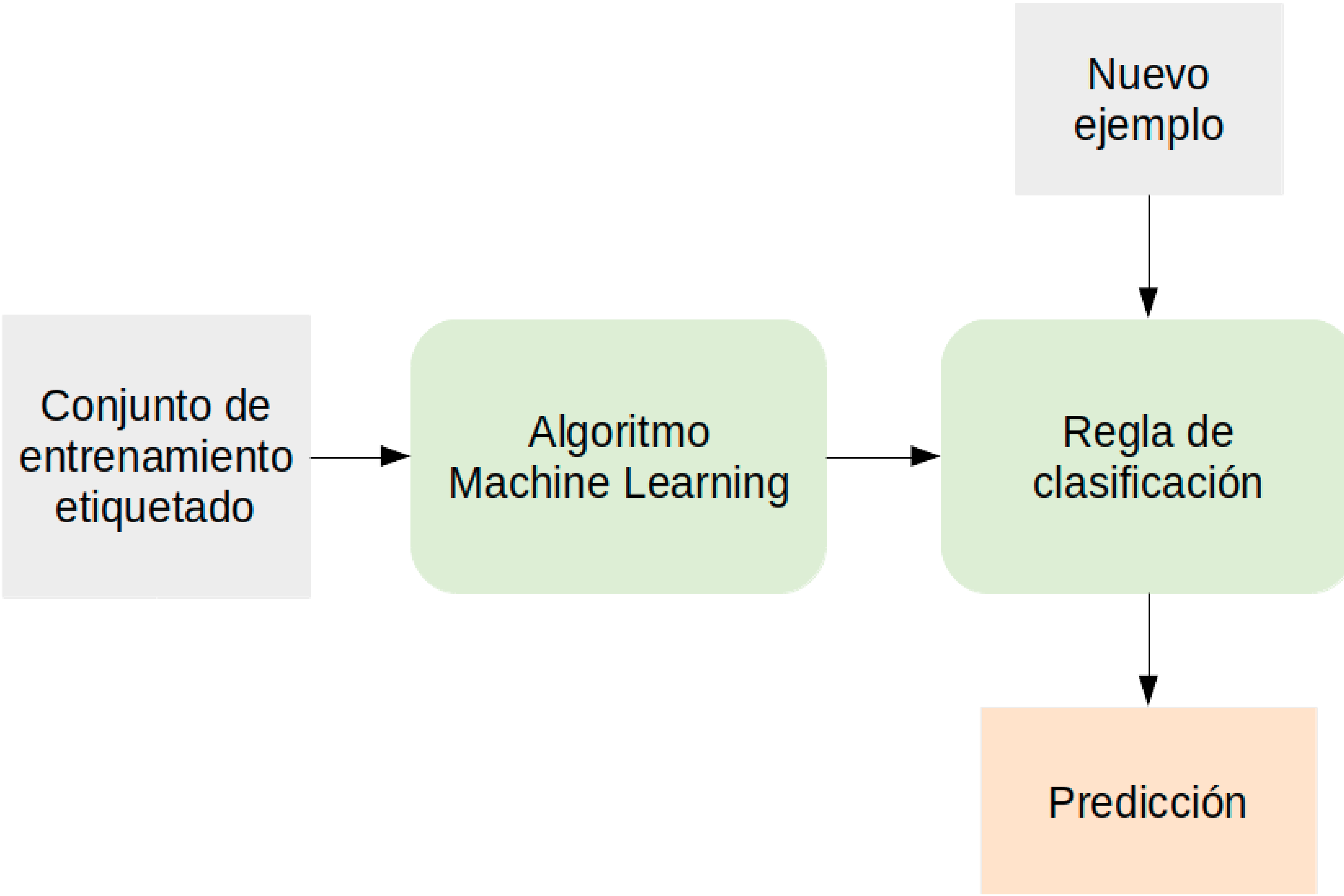
Machine Learning



Machine Learning



Clasificación



Clasificadores

- Random Forest
- Support Vector Machine

Decision Tree

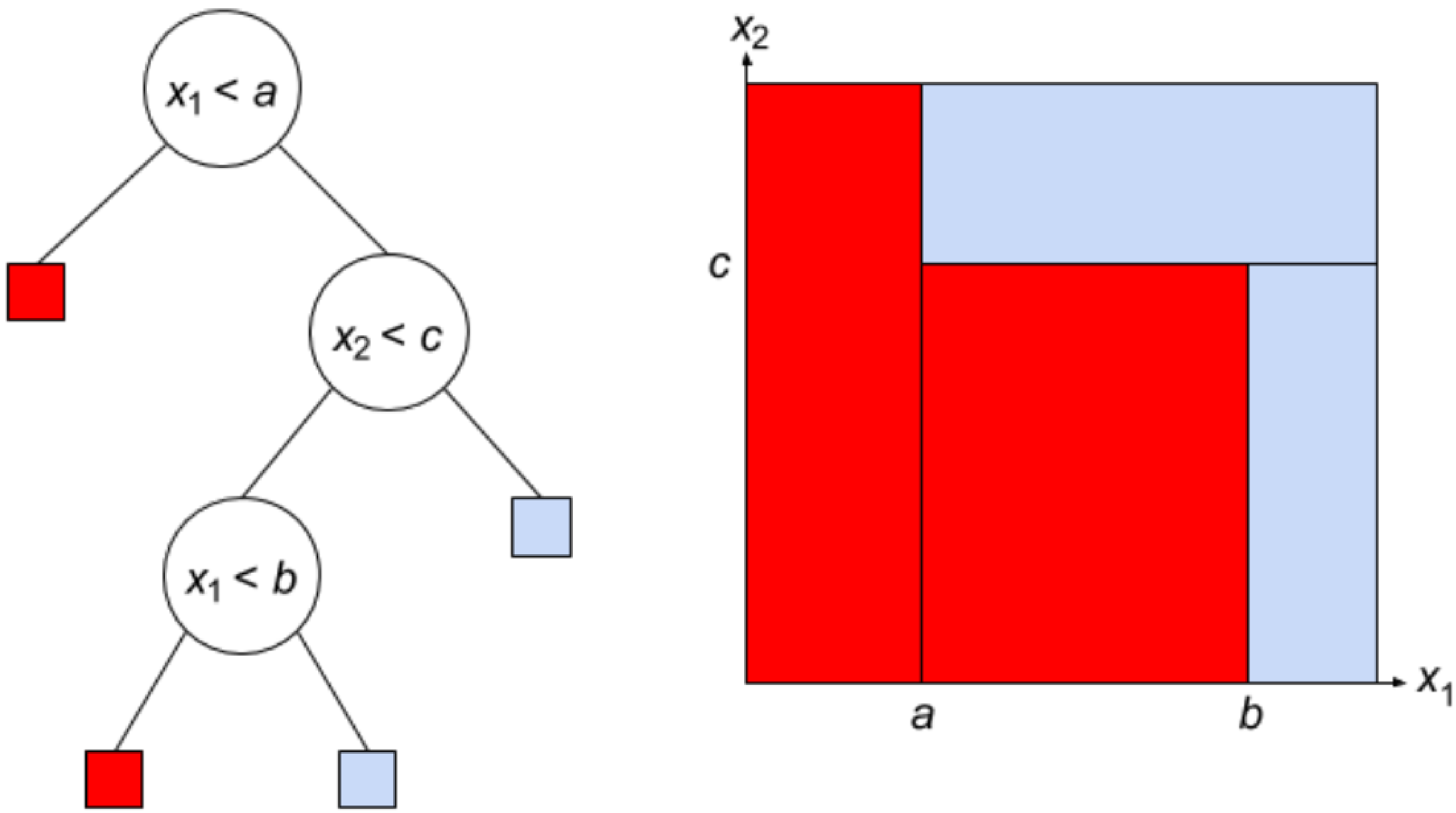


image: Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow by Aurélien Géron (O’Reilly). Copyright 2019 Aurélien Géron, 978-1-492-03264-9.

Random Forest

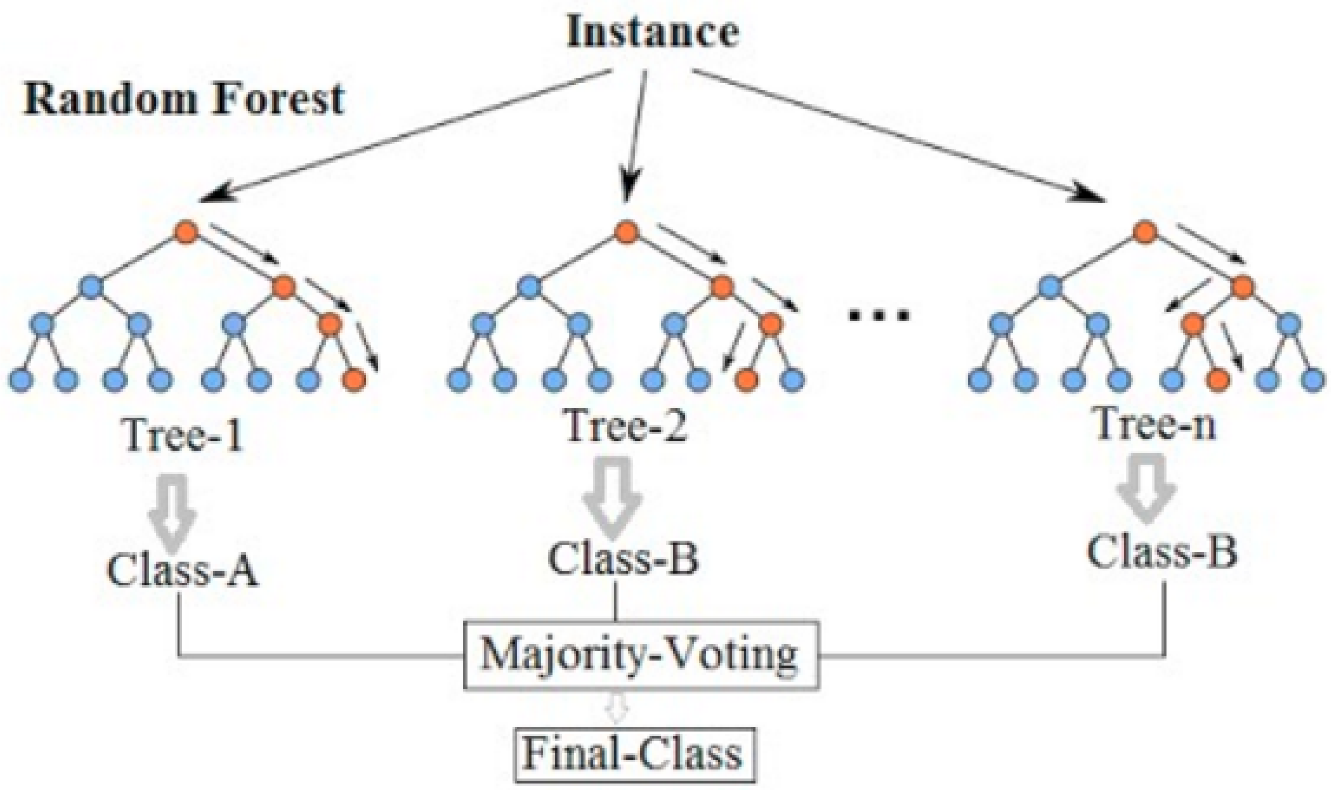
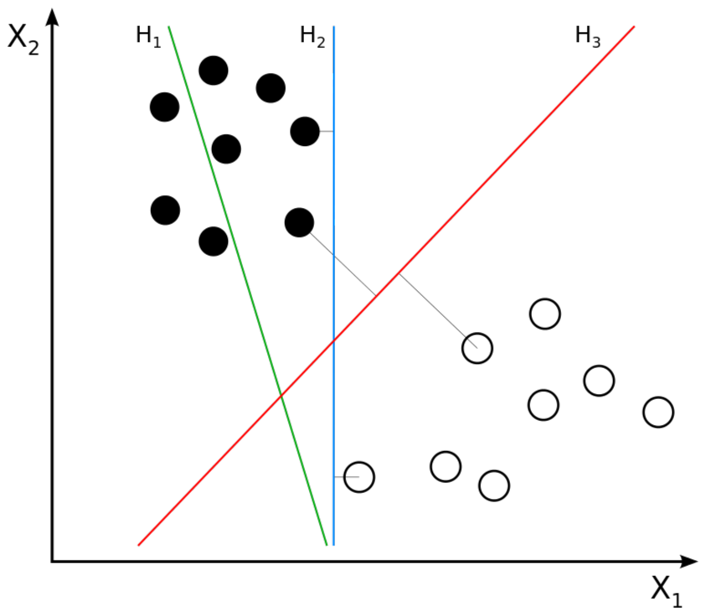
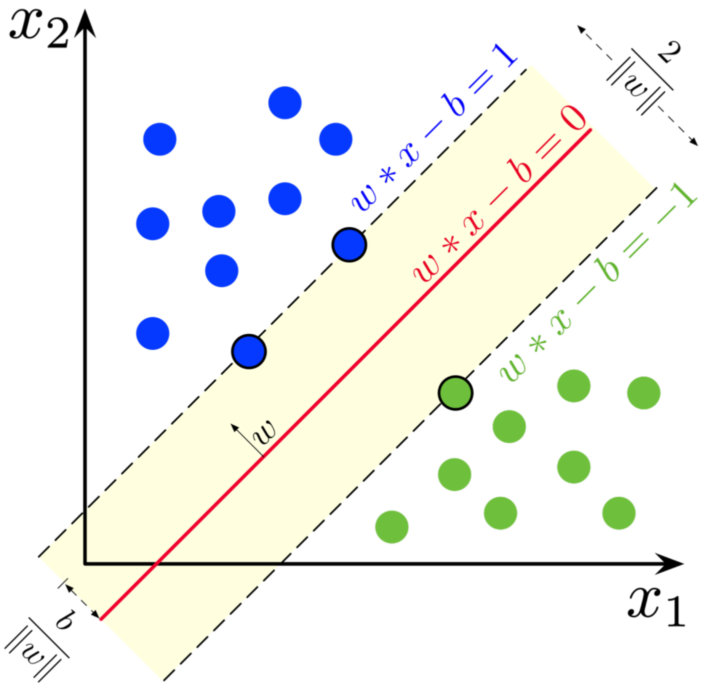


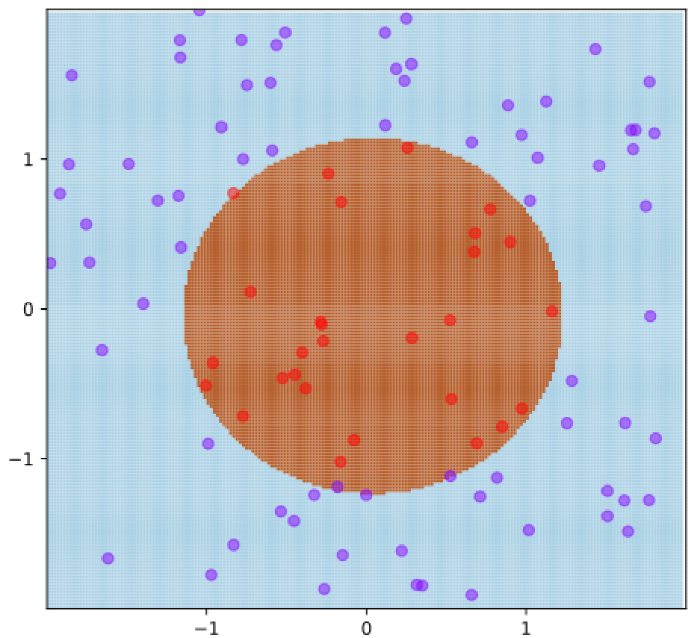
image: https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d

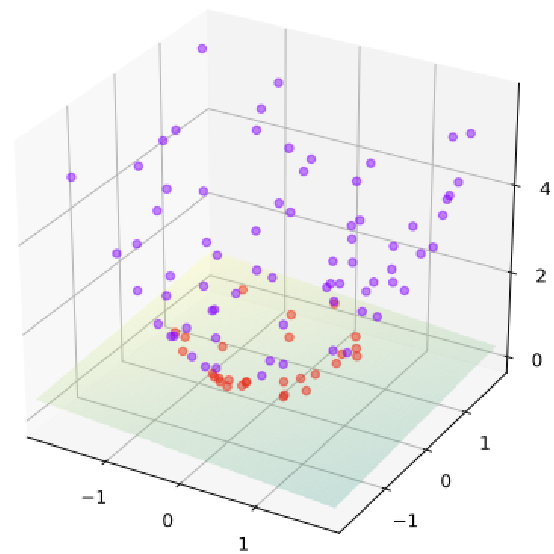
Support Vector Machine





Support Vector Machine





Support Vector Machine

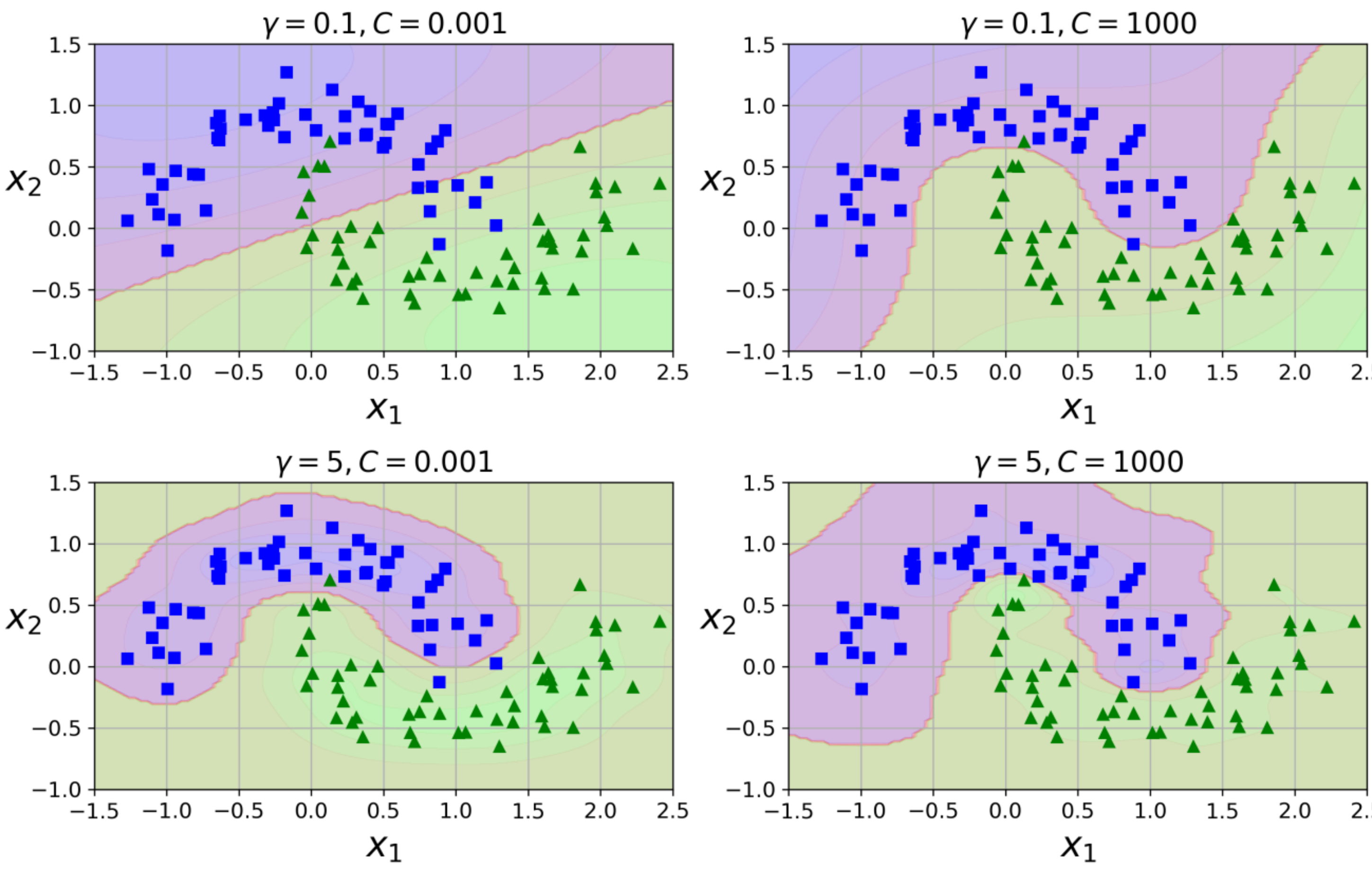


image: Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow by Aurélien Géron (O'Reilly). Copyright 2019 Aurélien Géron, 978-1-492-03264-9.

Support Vector Machine

- *Math*

$$\bar{w} \cdot \bar{u} + b = 0 \qquad Y_i \cdot \left(\bar{x}_i \cdot \bar{w} + b \right) - 1 = 0$$

$$min \frac{1}{2} \|\bar{w}\|^2 \qquad \sum_i \alpha_i \cdot Y_i = 0 \qquad \bar{w} = \sum_i \alpha_i \cdot Y_i \cdot \bar{x}_i$$

$$L = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j Y_i Y_j \bar{x}_i \bar{x}_j$$

$$K\left(x_i, x_j\right) = \phi\left(x_i\right) \cdot \phi\left(x_j\right) \qquad \left(\bar{x}_i \cdot \bar{x}_j + 1\right)^n \qquad e^{\frac{-\bar{x}_i - \bar{x}_j}{\sigma}}$$

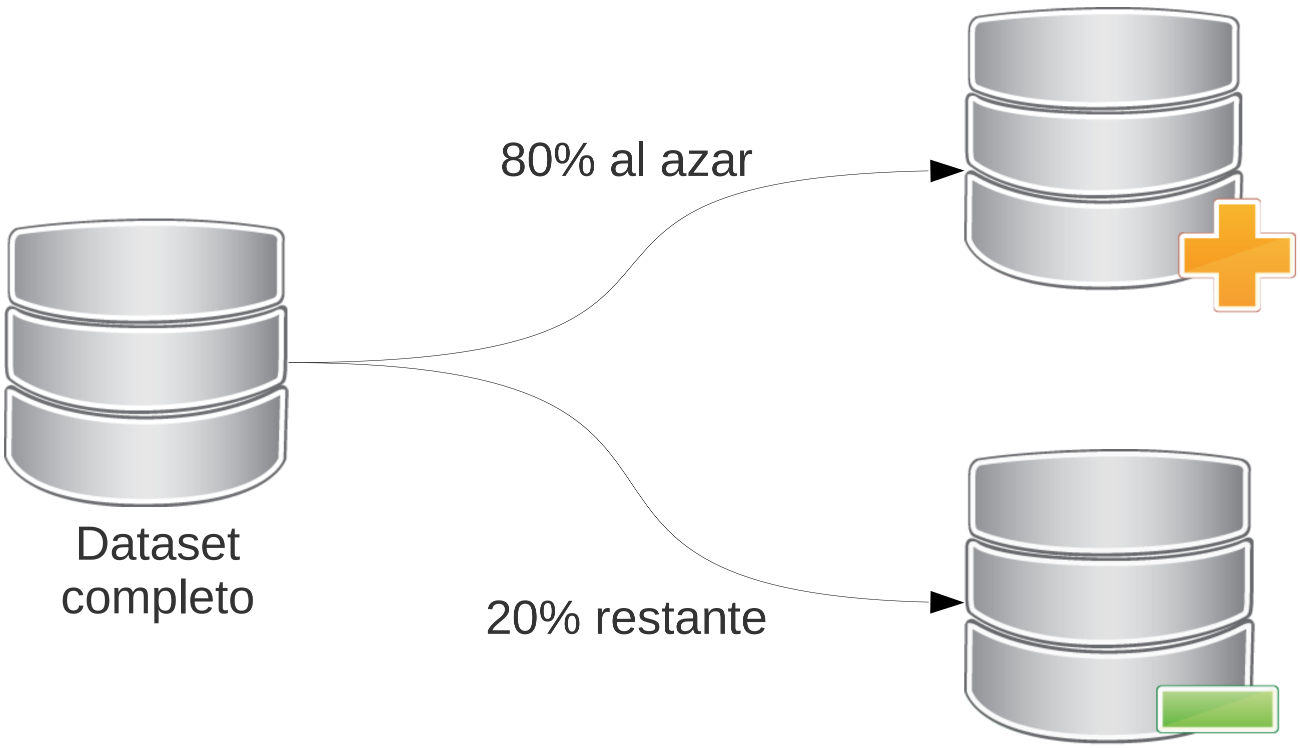
Características de la muestra

- **n total**: 256 trabajadores de la función pública.
- Edad media: 49,1± 8,2 (rango 26 a 67) años.
- SVI: 58,2%.
- Variables: ~80.

Características de la muestra

Variable		n	%
Sexo	Mujeres	166	**64,8**
	Hombres	90	35,2
Uso de ordenador	< 4h	54	21,1
	≥ 4h	202	**78,9**
Uso de lentes de contacto	Sí	112	43,7
	No	144	**56,3**
Ojo Seco	Sí	20	7,8
	No	236	**92,2**
Uso de lágrimas artificiales	Sí	38	14,8
	No	218	**85,2**

Dataset



Optimización de hiperparámetros

- **Scikit-learn**
  - Randomized Parameter Optimization.
  - Exhaustive Grid Search.

- **Population-Based Metaheuristics**
  - Evolutionary Computation.
  - Swarm Intelligence.

Optimización de hiperparámetros

- **Scikit-learn**
  - Randomized Parameter Optimization. *
  - Exhaustive Grid Search. *

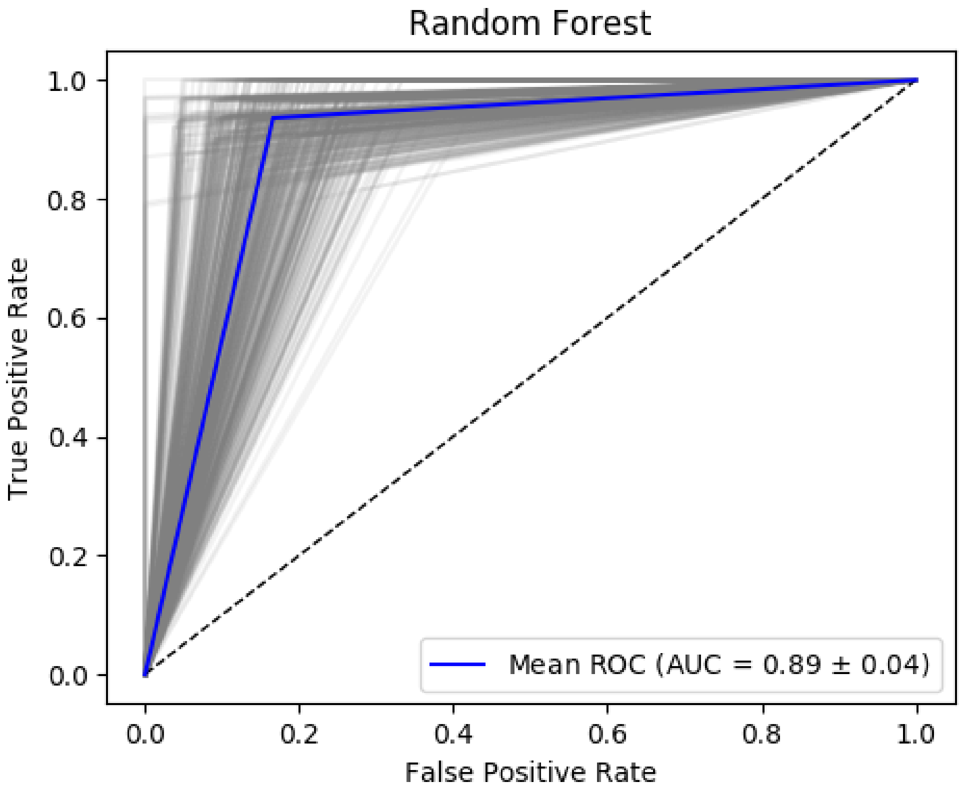
- **Population-Based Metaheuristics**
  - Evolutionary Computation.
  - Swarm Intelligence.

Herramientas

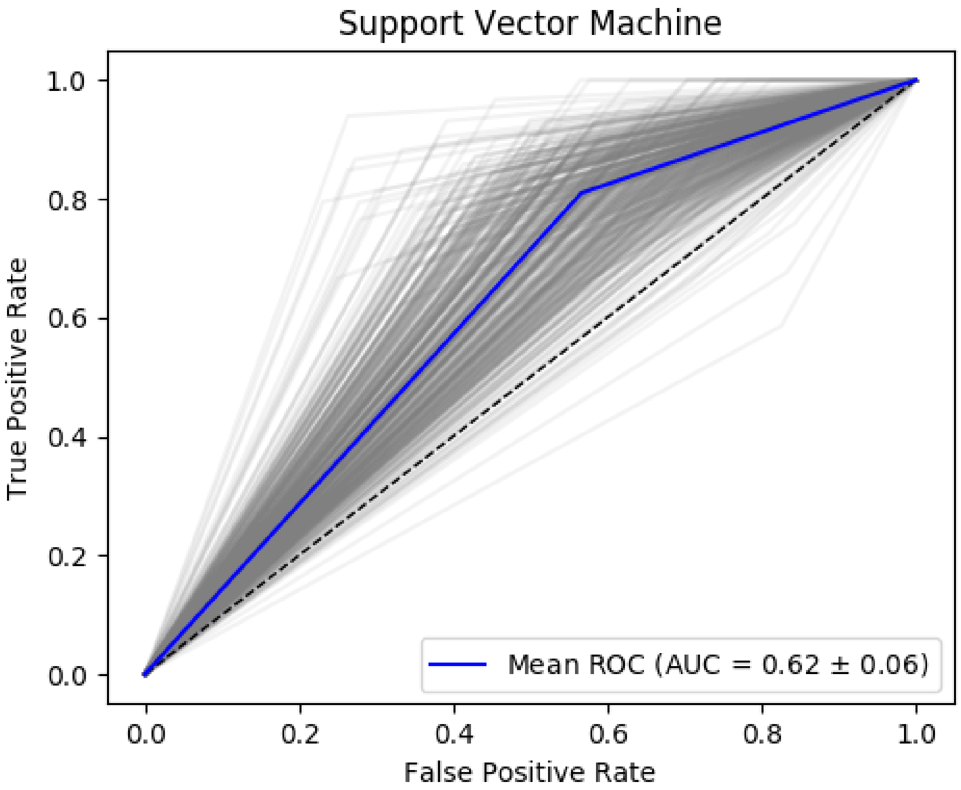
Repositorio

- GitHub:
  - https://github.com/rcrespocano/cvs-classifier/

Resultados

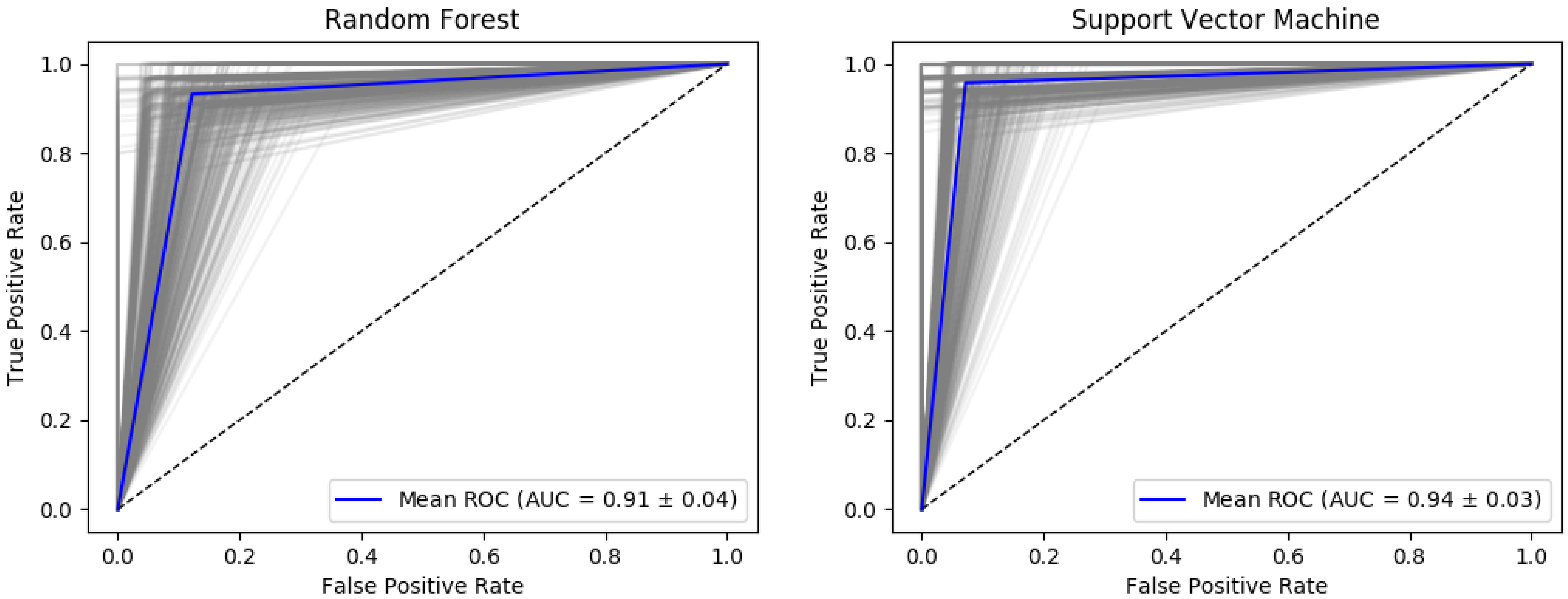


Random Forest
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
Mean ROC (AUC = 0.89 ± 0.04)



Support Vector Machine
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
Mean ROC (AUC = 0.62 ± 0.06)

Resultados


Random Forest
True Positive Rate
False Positive Rate
Mean ROC (AUC = 0.91 ± 0.04)
Support Vector Machine
True Positive Rate
False Positive Rate
Mean ROC (AUC = 0.94 ± 0.03)

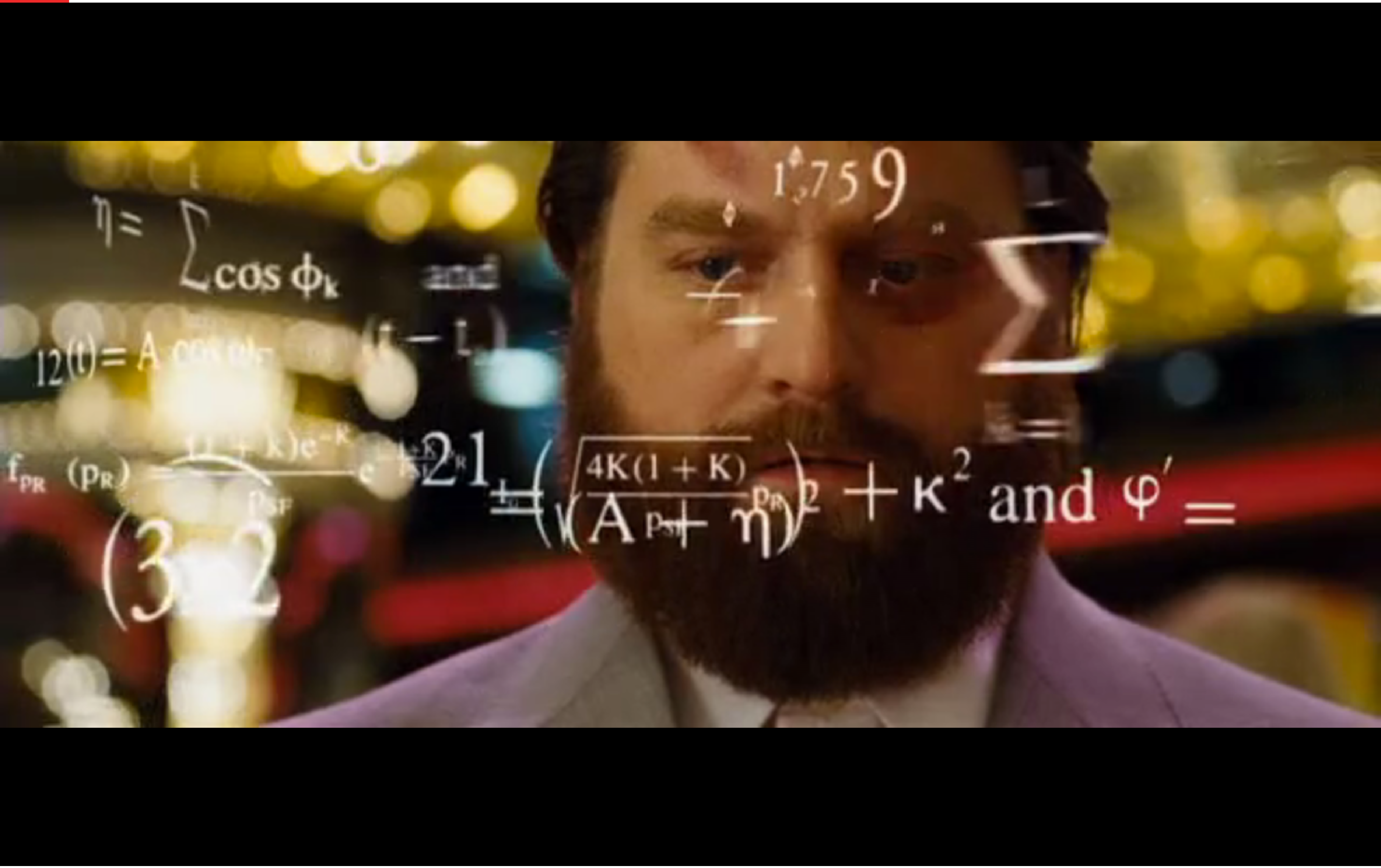
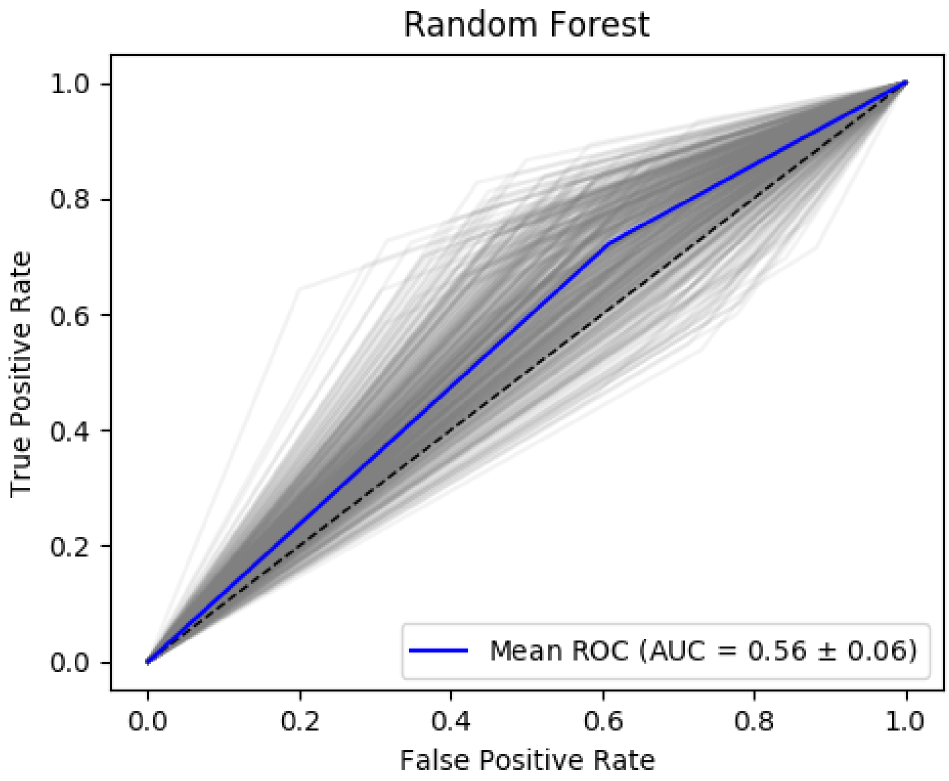
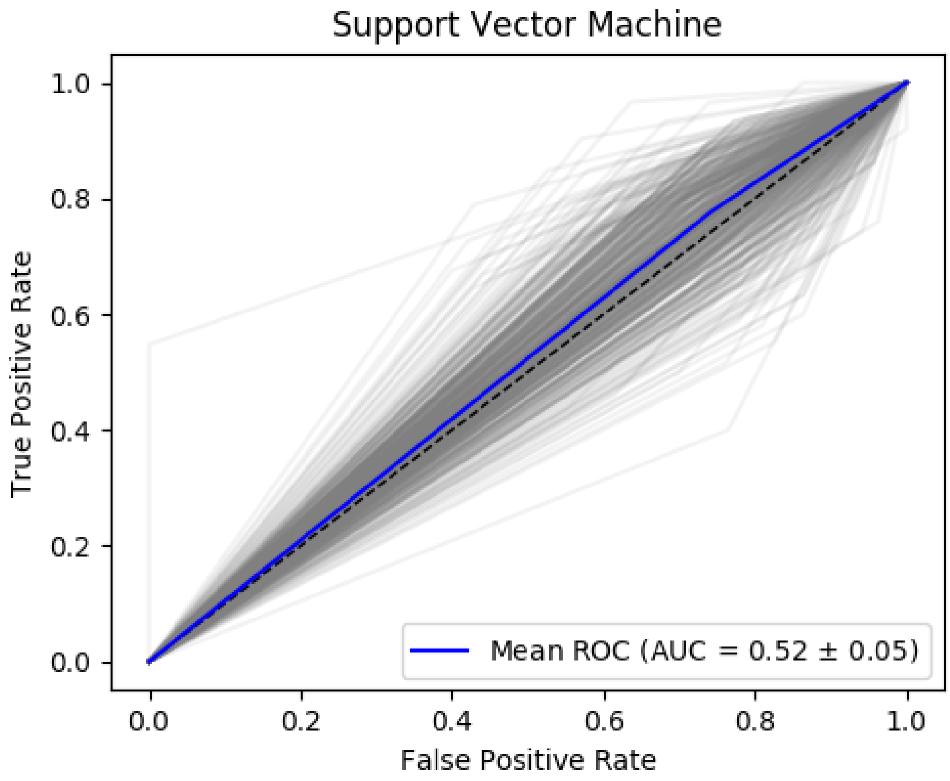


image: The Hangover; https://www.warnerbros.com/movies/hangover/

Resultados


Random Forest
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Mean ROC (AUC = 0.56 ± 0.06)


Support Vector Machine
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Mean ROC (AUC = 0.52 ± 0.05)

Hipótesis (I)

- Datos
  - Insuficiente cantidad de datos.
  - Training set no representativo.
  - Datos de baja calidad.
  - Características irrelevantes.
  - Overfitting / Underfitting.

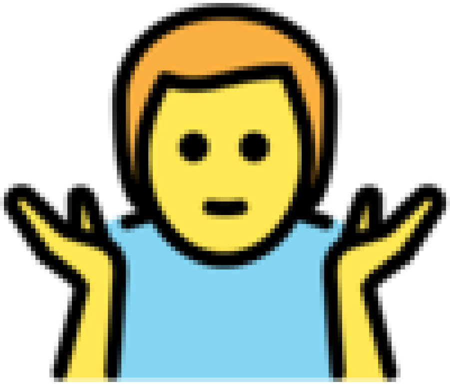
Hipótesis (II)

- No correlación entre síntomas y signos.
- No síndrome.
  - *“But in many cases, the difference between calling a condition a “syndrome” and simply describing a cluster of symptoms is more about language than about science or medicine [2].”*

    [2] https://www.health.harvard.edu/blog/times-changin-bringing-new-syndromes-201607019844

¿Podemos predecir el SVI?

- No lo sabemos (todavía).



¡Muchas gracias!

@anatauste_
@rcrespocano