



**UNIVERSIDAD NACIONAL DEL ALTIPLANO**  
Escuela Profesional del Ingeniería de Sistemas

## **LÓGICA PARA INTELIGENCIA ARTIFICIAL** **Árboles de Decisión**

Prof. Ing. José Luis Juárez Ruelas

Puno, Mayo del 2019

# Introducción

decision trees - Google Académico


201,022 Search Results - Keyword

← → ↺ 🏠

🔒 [https://www.sciencedirect.com/search?qs=decision tree&show=25&sortBy=relevance](https://www.sciencedirect.com/search?qs=decision+tree&show=25&sortBy=relevance)

⋮ 📄 🌐 🔊 🔍

Google Académico

 **ScienceDirect**

Journals & Books

?

Create account

Sign in

decision tree

Author name

Journal/book title

Volume

Issue

Pages

🔍

Advanced search

201,022 results

🔔 Set search alert

Refine by:

Years

☐ 2020 (24)

☐ 2019 (8,871)

☐ 2018 (15,630)

Show more

Article type

☐ Review articles (11,504)

☐ Research articles (139,109)

☐ Encyclopedia (2,506)

☐ Book chapters (21,695)

Show more

Publication title

☐ Expert Systems with Applications (3,298)

☐ European Journal of Operational Research (2,994)

☐ Forest Ecology and Management (2,724)

Show more

☐ Download selected articles

📄 Export

sorted by *relevance* | *date*

☐ Research article • Full text access

**Decision tree** classifiers for evidential attribute values and class labels

Fuzzy Sets and Systems, Volume 366, 1 July 2019, Pages 46-62

Asma Trabelsi, Zied Elouedi, Eric Lefevre

📄 Download PDF

Abstract

Export

☐ Research article • Full text access

**An assessment of the risk factors for vitamin D deficiency using a decision tree model**

Diabetes & Metabolic Syndrome: Clinical Research & Reviews, Volume 13, Issue 3, May-June 2019, Pages 1773-1777

Kayhan Gonoodi, Maryam Tayefi, Maryam Saberi-Karimian, Alireza Amirabadi zadeh, ... Majid Ghayour Mobarhan

📄 Download PDF

Abstract

Export

Want a richer search experience?

Sign in for personalized recommendations, search alerts, and [more](#).

Sign in

☐ Research article • Full text access

**Dispersion Ratio based Decision Tree Model for Classification**

Expert Systems with Applications, Volume 116, February 2019, Pages 1-9

Smita Roy, Samrat Mondal, Asif Ekbal, Maunendra Sankar Desarkar

📄 Download PDF

Abstract

Export

☐ Research article • Full text access

Feedback

🗨

2

# Introdução

The screenshot shows a Google Scholar search for "decision trees". The browser tabs include "decision trees - Google Acadêm" and "201,022 Search Results - Keywo". The search bar shows "decision trees" with a magnifying glass icon. The results are sorted by "Aproximadamente 2.310.000 resultados (0,04 s)". On the left, there are filters for "Artigos", "A qualquer momento" (Desde 2019, Desde 2018, Desde 2015, Período específico...), "Classificar por relevância" (Classificar por data), "Em qualquer idioma" (Pesquisar páginas em Português), and checkboxes for "incluir patentes" and "incluir citações", along with a "Criar alerta" button. The main results list includes:

- Induction of decision trees** by JR Quinlan - Machine learning, 1986 - Springer. The technology for building knowledge-based systems by inductive inference from examples has been demonstrated successfully in several practical applications. This paper summarizes an approach to synthesizing **decision trees** that has been used in a variety of ... Cited by 20139. [PDF] springer.com
- Simplifying decision trees** by JR Quinlan - International journal of man-machine studies, 1987 - Elsevier. Many systems have been developed for constructing **decision trees** from collections of examples. Although the **decision trees** generated by these methods are accurate and efficient, they often suffer the disadvantage of excessive complexity and are therefore ... Cited by 2318. [PDF] mit.edu
- Part-of-speech tagging using decision trees** by L Márquez, H Rodríguez - European Conference on Machine Learning, 1998 - Springer. We have applied inductive learning of statistical **decision trees** to the Natural Language Processing (NLP) task of morphosyntactic disambiguation (Part Of Speech Tagging). Previous work showed that the acquired language models are independent enough to be ... Cited by 3639. [PDF] springer.com
- Constructing optimal binary decision trees is NP-complete** by H Laurent, RL Rivest - Information processing letters, 1976 - people.csail.mit.edu. We demonstrate that constructing optimal binary **decision trees** is an NP-complete problem, where an optimal tree is one which minimizes the expected number of tests required to identify the unknown object. Precise definitions of NP-complete problems are given in ... Cited by 962. [PDF] mit.edu

Below the results, there is a section "Pesquisas relacionadas" (Related searches) with links to:

- decision trees induction
- fuzzy decision trees
- data mining decision trees
- decision trees classification
- decision trees neural networks
- decision trees machine learning
- boosted decision trees
- quinlan decision trees

At the bottom, there is a link for "Induction of fuzzy decision trees".

# Introducción

The screenshot shows a web browser with multiple tabs. The active tab is 'Search | arXiv e-print repository'. The address bar shows the URL 'https://arxiv.org/search/?query=decision+tree&searchtype=all'. The page header includes the Cornell University logo and the arXiv logo. A search bar at the top right contains the text 'decision tree'. Below the search bar, it says 'Showing 1–50 of 1,422 results for all: decision tree'. There are buttons for 'Search v0.5 released 2018-12-20', 'Feedback?', 'All fields', and 'Search'. Below the search bar, there are radio buttons for 'Show abstracts' (selected) and 'Hide abstracts'. There is a link for 'Advanced Search'. Below the search bar, there is a section for '50 results per page. Sort results by Announcement date (newest first) Go'. There are pagination buttons for 1, 2, 3, 4, 5, and a 'Next' button. The first result is '1. arXiv:1905.00355 [pdf, other] econ.TH Compactification of Extensive Form Games and Belief in the Opponents' Future Rationality'. The authors are 'Shuige Liu'. The abstract is 'We introduce an operation, called compactification, to reduce an extensive form to a compact one where each decision node in the game tree can be assigned to more than one player. Motivated by Thompson (1952)'s interchange of decision nodes, we attempt to capture the notion o...'. The second result is '2. arXiv:1905.00328 [pdf, other] cs.LG cs.AI stat.ML Interpretable multiclass classification by MDL-based rule lists'. The authors are 'Hugo M. Proença, Matthijs van Leeuwen'. The abstract is '...mining community because they are inherently easier to understand and explain than their more complex counterparts. Examples of interpretable classification models include decision trees, rule sets, and rule lists. Learning such models often involves optimizing hyperparameters, which typically requires substantial amou...'. The page number '4' is visible in the bottom right corner.

decision trees - Google Académico X 201,022 Search Results - Keyword X Search | arXiv e-print repository X +

Google Académico

Cornell University

arXiv

We gratefully acknowledge support from the Simons Foundation and member institutions.

Search... All fields Search

Help | Advanced Search Login

Showing 1–50 of 1,422 results for all: decision tree

Search v0.5 released 2018-12-20 Feedback?

decision tree All fields Search

☒ Show abstracts ☐ Hide abstracts

Advanced Search

50 results per page. Sort results by Announcement date (newest first) Go

1 2 3 4 5 ... Next

1. arXiv:1905.00355 [pdf, other] econ.TH

**Compactification of Extensive Form Games and Belief in the Opponents' Future Rationality**

Authors: Shuige Liu

**Abstract:** We introduce an operation, called compactification, to reduce an extensive form to a compact one where each decision node in the game tree can be assigned to more than one player. Motivated by Thompson (1952)'s interchange of decision nodes, we attempt to capture the notion o... [▼ More](#)

Submitted 1 May, 2019; originally announced May 2019.

2. arXiv:1905.00328 [pdf, other] cs.LG cs.AI stat.ML







**Interpretable multiclass classification by MDL-based rule lists**

Authors: Hugo M. Proença, Matthijs van Leeuwen

**Abstract:** ...mining community because they are inherently easier to understand and explain than their more complex counterparts. Examples of interpretable classification models include decision trees, rule sets, and rule lists. Learning such models often involves optimizing hyperparameters, which typically requires substantial amou... [▼ More](#)

Submitted 1 May, 2019; originally announced May 2019.







# Introducción

Gender	Occupation	App
F	Study	Pokemon Go 
F	Work	WhatsApp 
M	Work	Snapchat 
F	Work	WhatsApp 
M	Study	Pokemon Go 
M	Study	Pokemon Go 


## ¿Cual aplicación recomendamos ?

- Para una mujer que trabaja en oficina
- Para un hombre que trabaja en fabrica
- Para alguien que va al colegio

# Introducción







Gender	Occupation	App
F	Study	Pokemon Go 
F	Work	WhatsApp 
M	Work	Snapchat 
F	Work	WhatsApp 
M	Study	Pokemon Go 
M	Study	Pokemon Go 

# Introducción

Gender	Occupation	App	
F	Study	Pokemon Go	
F	Work	WhatsApp	
M	Work	Snapchat	
F	Work	WhatsApp	
M	Study	Pokemon Go	
M	Study	Pokemon Go	



# Introducción







Gender	Occupation	App
F	Study	Pokemon Go 
F	Work	WhatsApp 
M	Work	Snapchat 
F	Work	WhatsApp 
M	Study	Pokemon Go 
M	Study	Pokemon Go 

Entre **género** y **ocupación**

¿Cuál variable le parece más decisivo para predecir la aplicación descargarán los usuarios?



# Introducción

Gender	Occupation	App
F	Study	Pokemon Go 
F	Work	WhatsApp 
M	Work	Snapchat 
F	Work	WhatsApp 
M	Study	Pokemon Go 
M	Study	Pokemon Go 

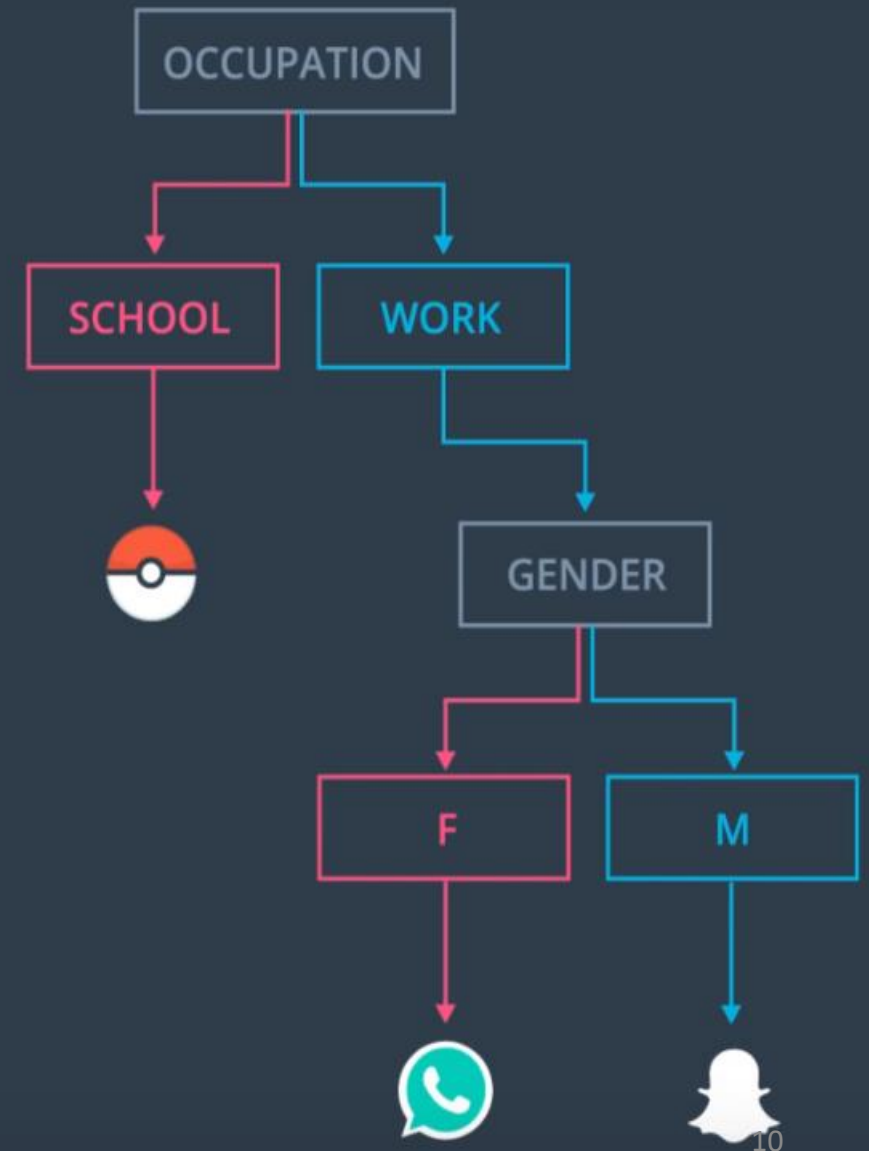
Entre **género** y **ocupación**

¿Cuál variable le parece más decisivo para predecir qué aplicación descargarán los usuarios?

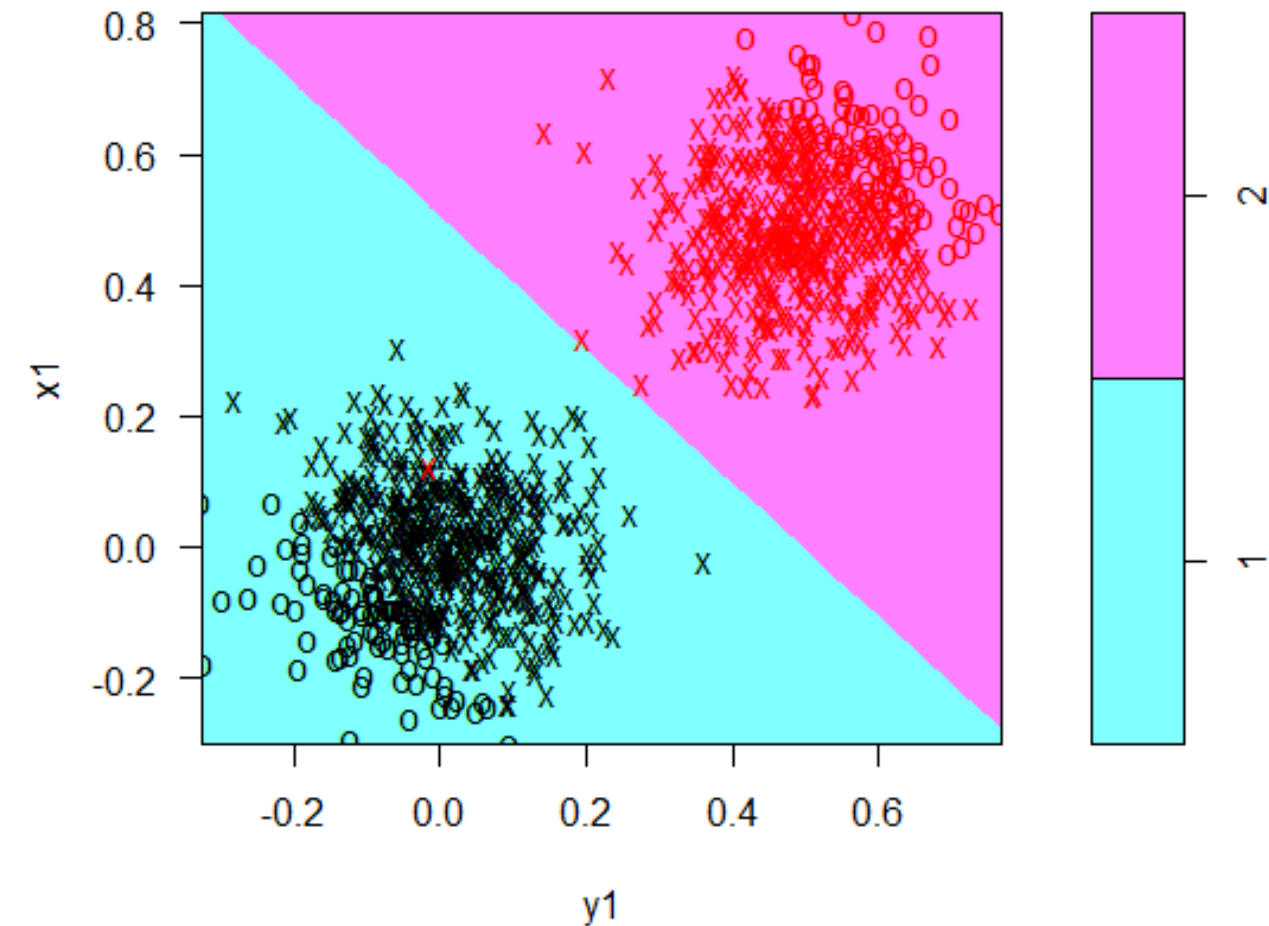
# Introducción

## Recommending Apps

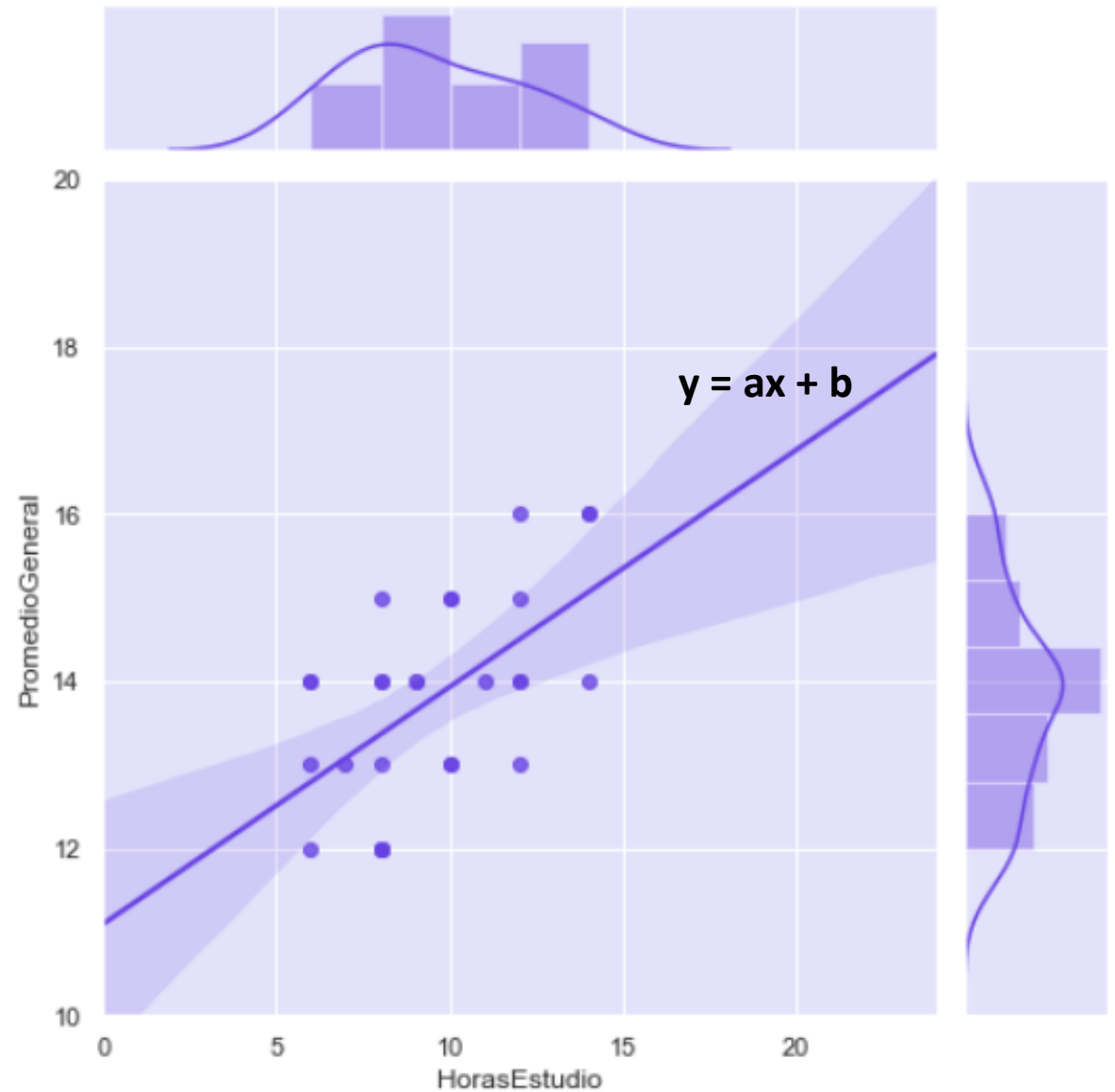
Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	



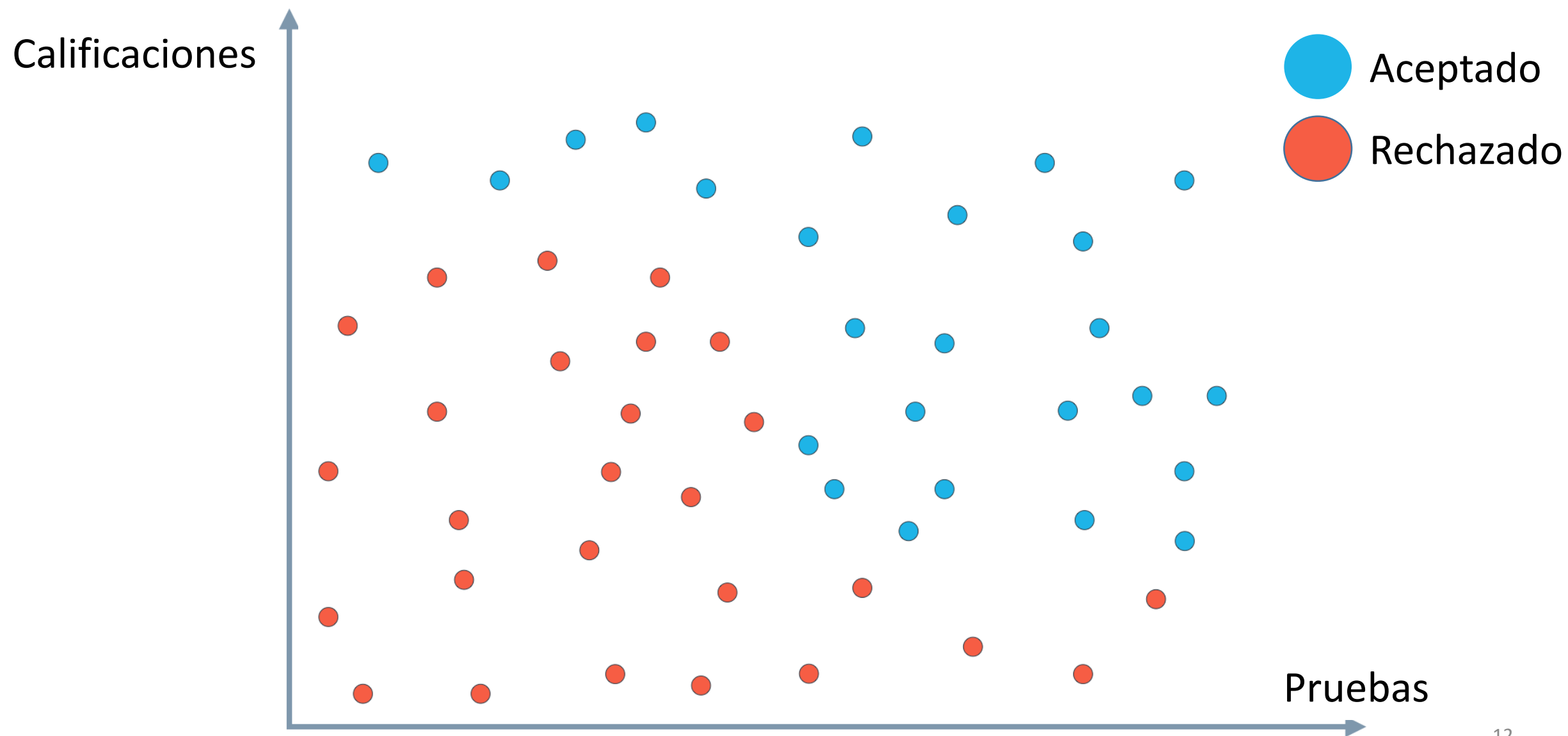
# Clasificación



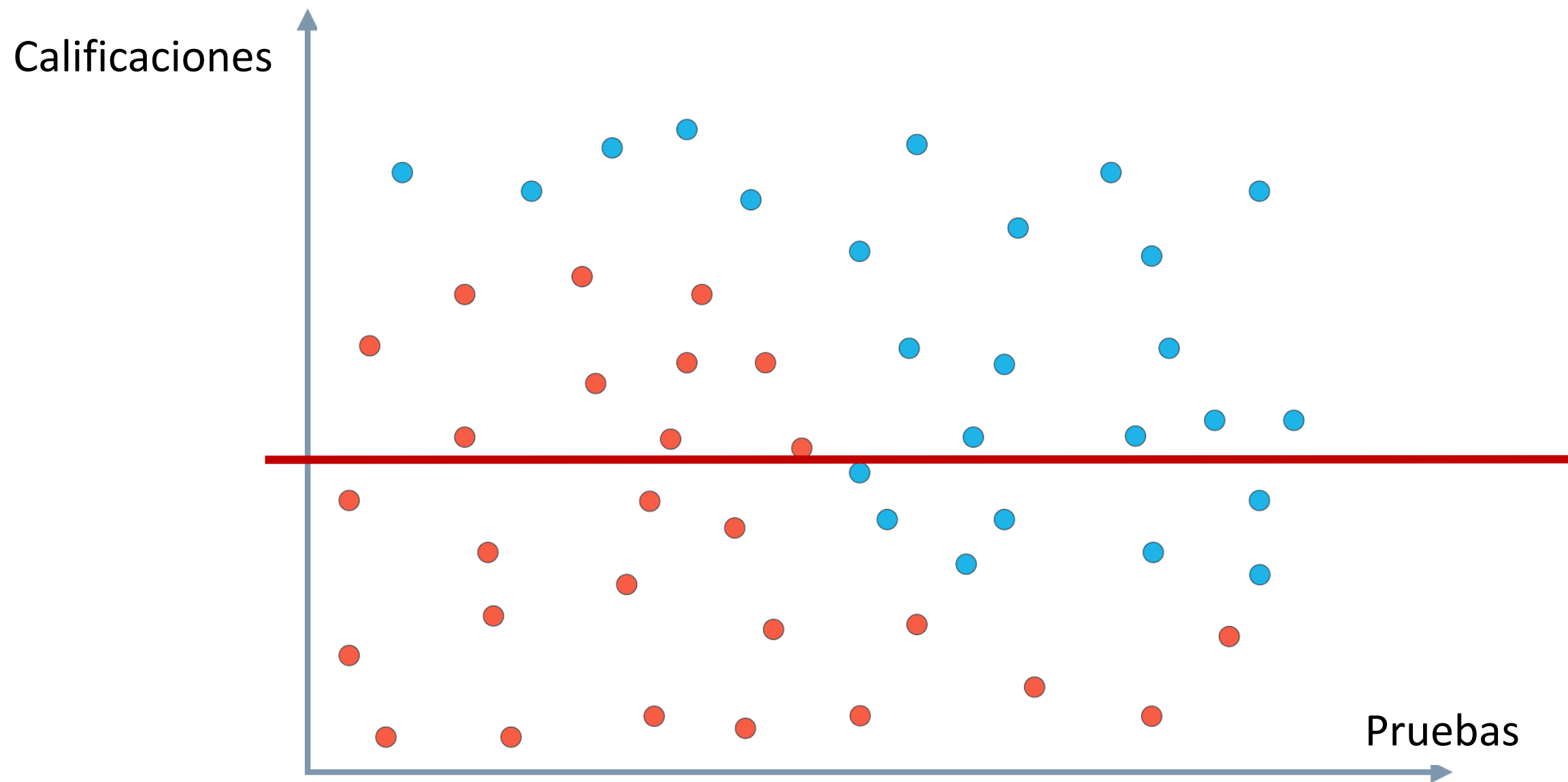
# Regresión



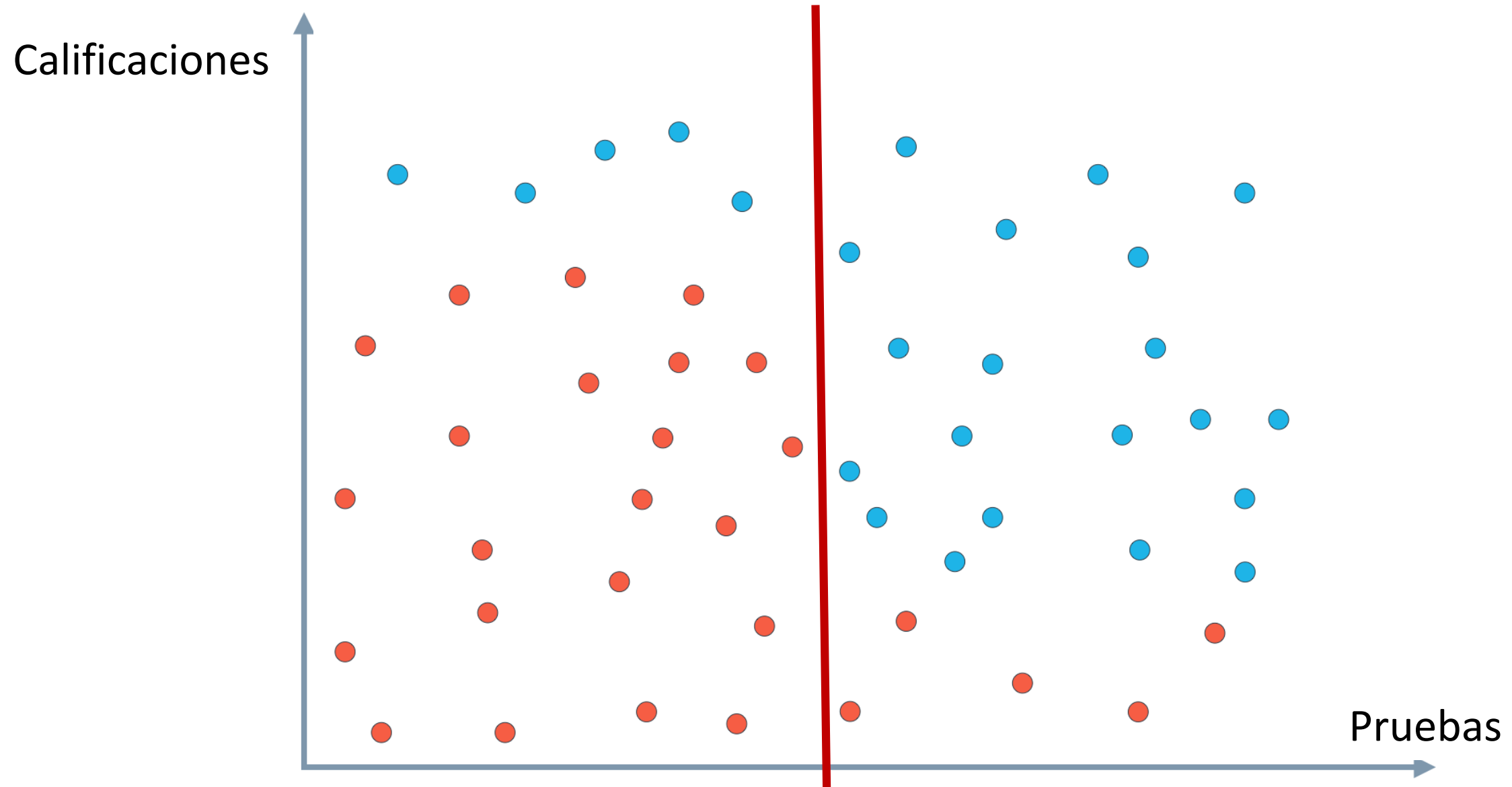
# La mejor recta de división para ser aceptado en un programa de estudios



# ¿Cuál es la mejor recta de división?

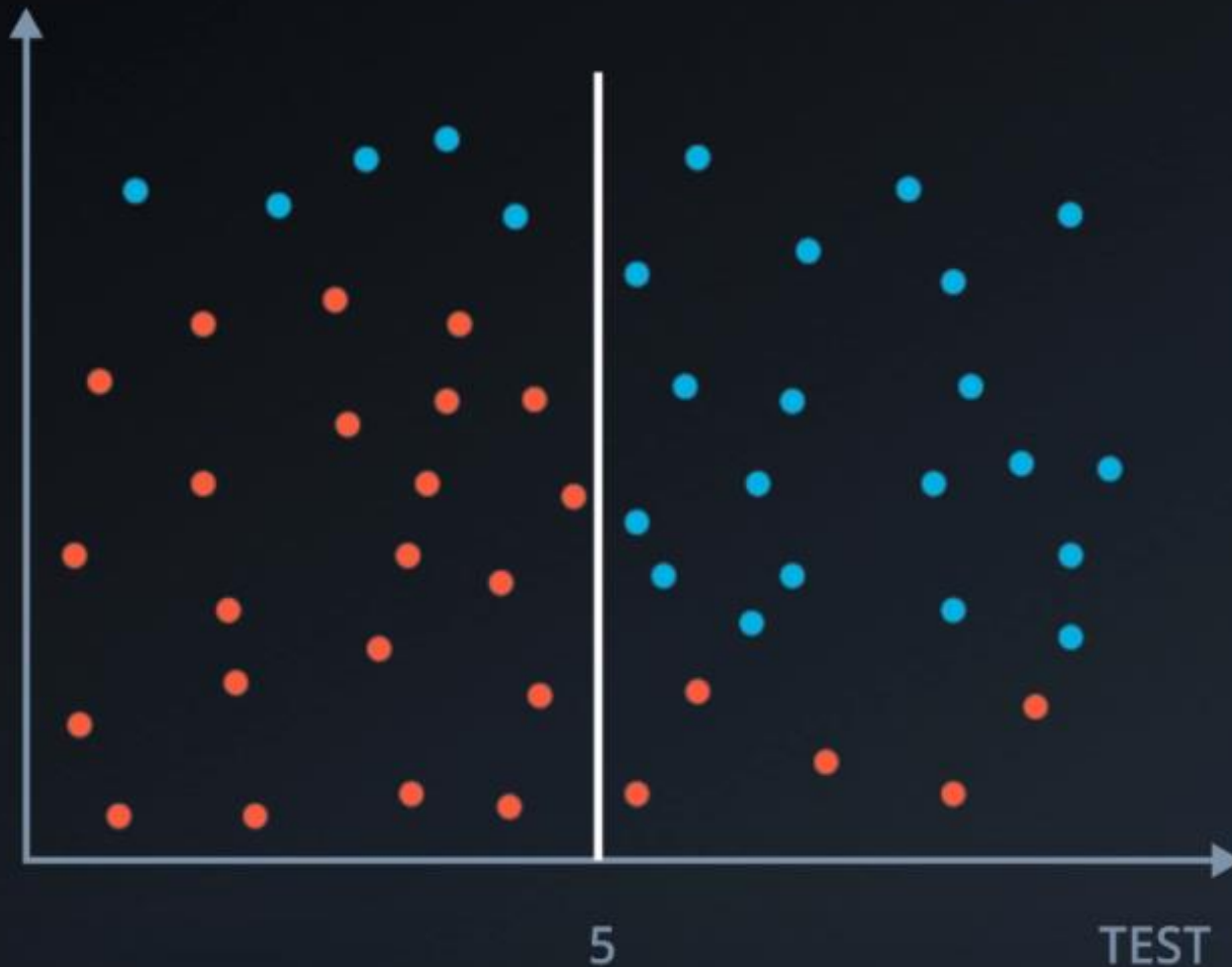


# La mejor recta de división



# Creando el árbol de decisión

GRADES

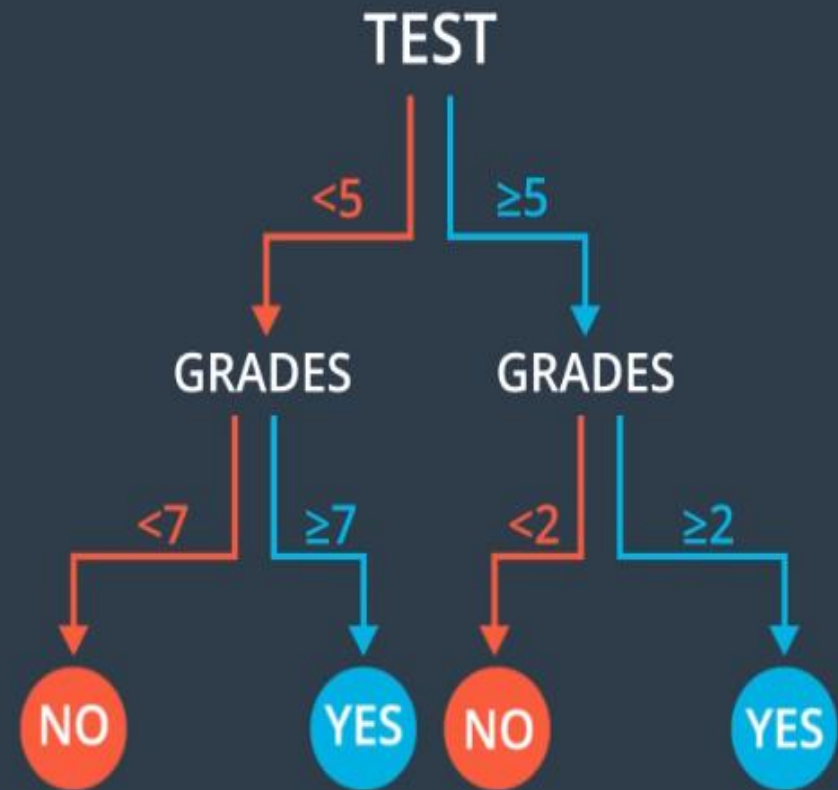


TEST





# Creando el árbol de decisión



# Entropía

## Entropy



Solido

Entropía baja



Líquido

Entropía media

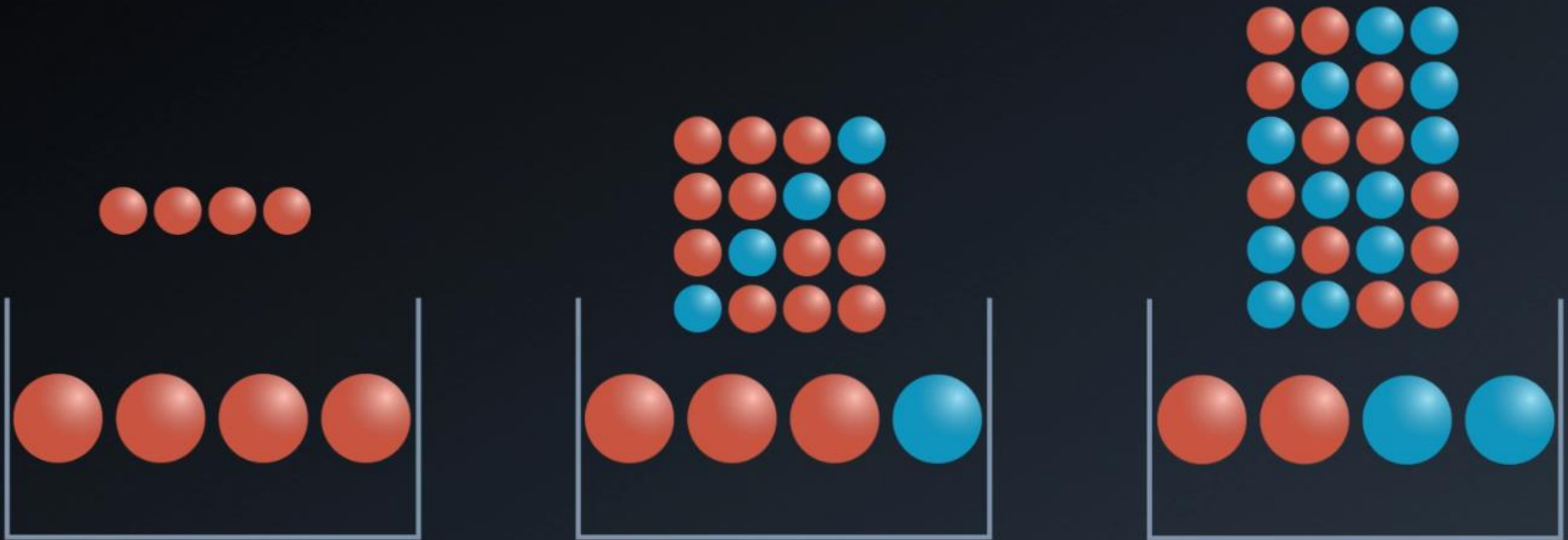


Vapor

Entropía alta

# Entropía

Def. En termodinámica, la entropía (simbolizada como  $S$ ) es una magnitud física para un sistema termodinámico en equilibrio. Mide el número de microestados compatibles con el macroestado de equilibrio, también se puede decir que mide el grado de organización del sistema.



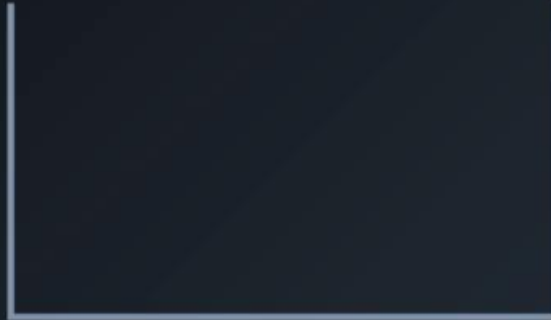
Entropía baja

Entropía media

Entropía alta

# Entropía

Game



# Entropía

## Game






Win lots of money!



No money

# Entropía

## Probability of Winning

	$P_{(red)}$	$P_{(blue)}$	$P_{(winning)}$
	1	1	$1 \times 1 \times 1 \times 1 = 1$
	0.75	0.25	$0.75 \times 0.75 \times 0.75 \times 0.25 = 0.105$
	0.5	0.5	$0.5 \times 0.5 \times 0.5 \times 0.5 = 0.0625$

¿Qué función nos ayudará a convertir los productos en sumas?

## Products

$$0.75 * 0.75 * 0.75 * 0.25 = 0.105$$



- Sin
- Cos
- Log
- Exp



# ¿Qué función nos ayudará a convertir los productos en sumas?

$$0.75 * 0.75 * 0.75 * 0.25 = 0.105$$




$$\log(0.75) + \log(0.75) + \log(0.75) + \log(0.25) = -3.245$$



$$\text{Log}(ab) = \text{Log}(a) + \text{Log}(b)$$

# $-Log_2$ , promedio de entropía

## Entropy

	$P$ (red)	$P$ (blue)	$P$ (winning)	$-\log_2$ ( $P(\text{winning})$ )	Entropy (average)
	1	1	$1 \times 1 \times 1 \times 1 = 1$	$0 + 0 + 0 + 0 = 0$	0
	0.75	0.25	$0.75 \times 0.75 \times 0.75 \times 0.25 = 0.105$	$0.415 + 0.415 + 0.415 + 2 = 3.245$	0.81
	0.5	0.5	$0.5 \times 0.5 \times 0.5 \times 0.5 = 0.0625$	$1 + 1 + 1 + 1 = 4$	1

# Resumen

## Entropy



$$Entropy = -\frac{5}{8}\log_2\left(\frac{5}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) = 0.9544$$

# Ecuación: Entropía

## Entropy



$$Entropy = -\frac{m}{m+n} \log_2 \left( \frac{m}{m+n} \right) - \frac{n}{m+n} \log_2 \left( \frac{n}{m+n} \right)$$

# Entropía multiclase

$$entropy = -\frac{m}{m+n} \log_2\left(\frac{m}{m+n}\right) - \frac{n}{m+n} \log_2\left(\frac{n}{m+n}\right)$$

$$p_1 = \frac{m}{m+n} \qquad p_2 = \frac{n}{m+n}$$

$$entropy = -p_1 \log_2(p_1) - p_2 \log_2(p_2)$$

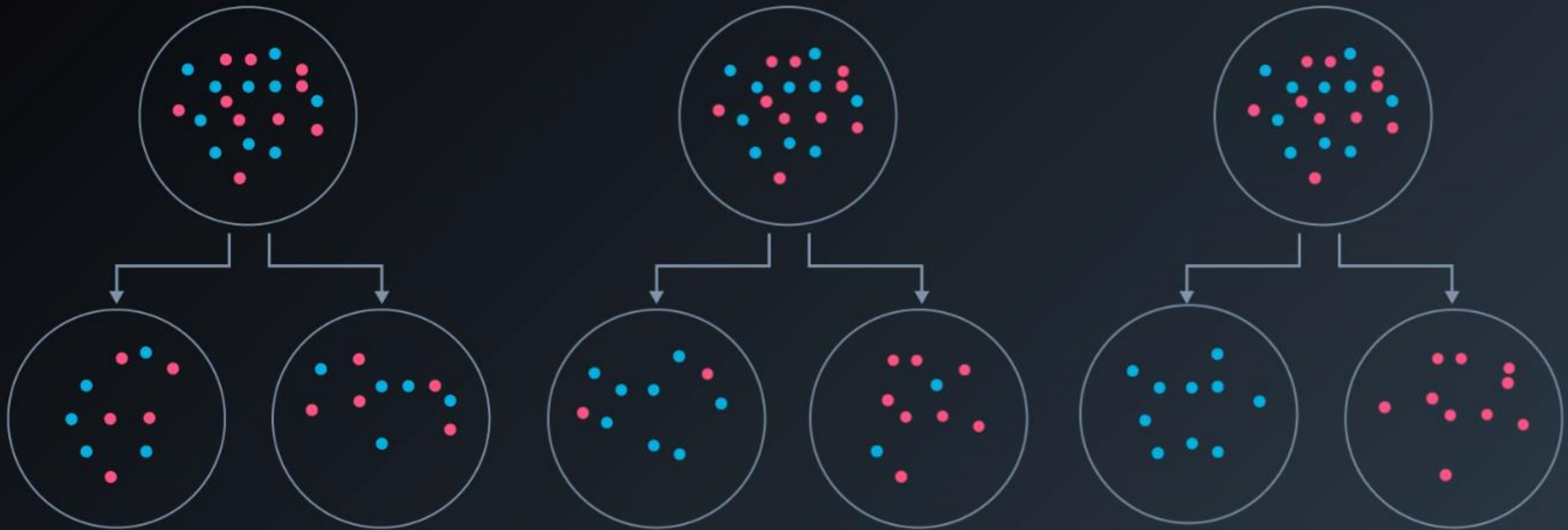
Esta ecuación de entropía se puede extender al caso de varias clases, donde tenemos tres o más valores posibles:

$$entropy = -p_1 \log_2(p_1) - p_2 \log_2(p_2) - \dots - p_n \log_2(p_n)$$

$$entropy = -\sum_{i=1}^n p_i \log_2(p_i)$$

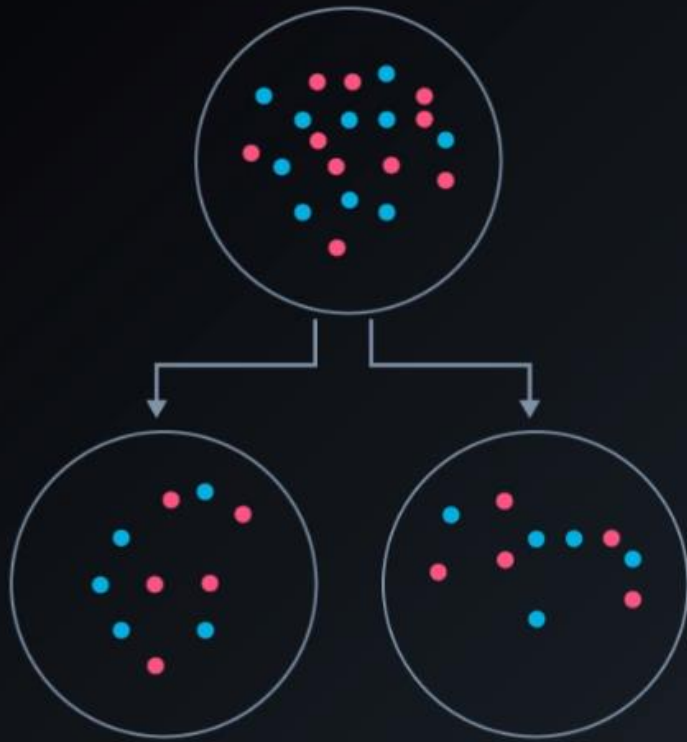
# Ganancia de información (Information Gain)

## Information Gain

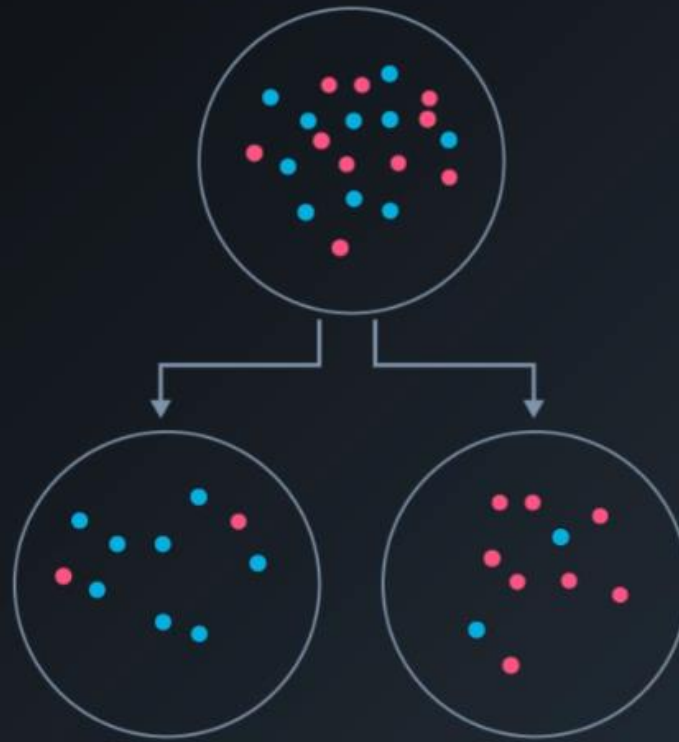




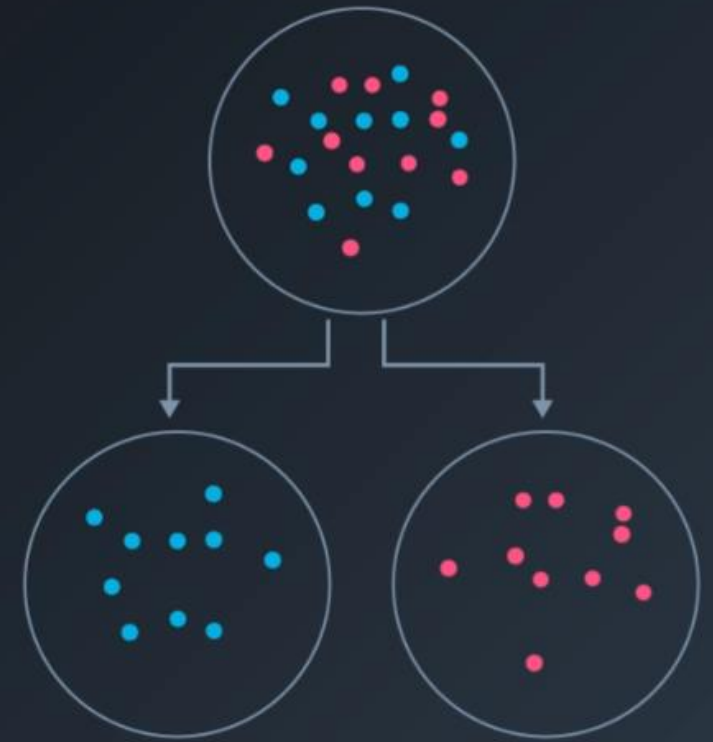
# Ganancia de información (Information Gain)



Information gain = 0



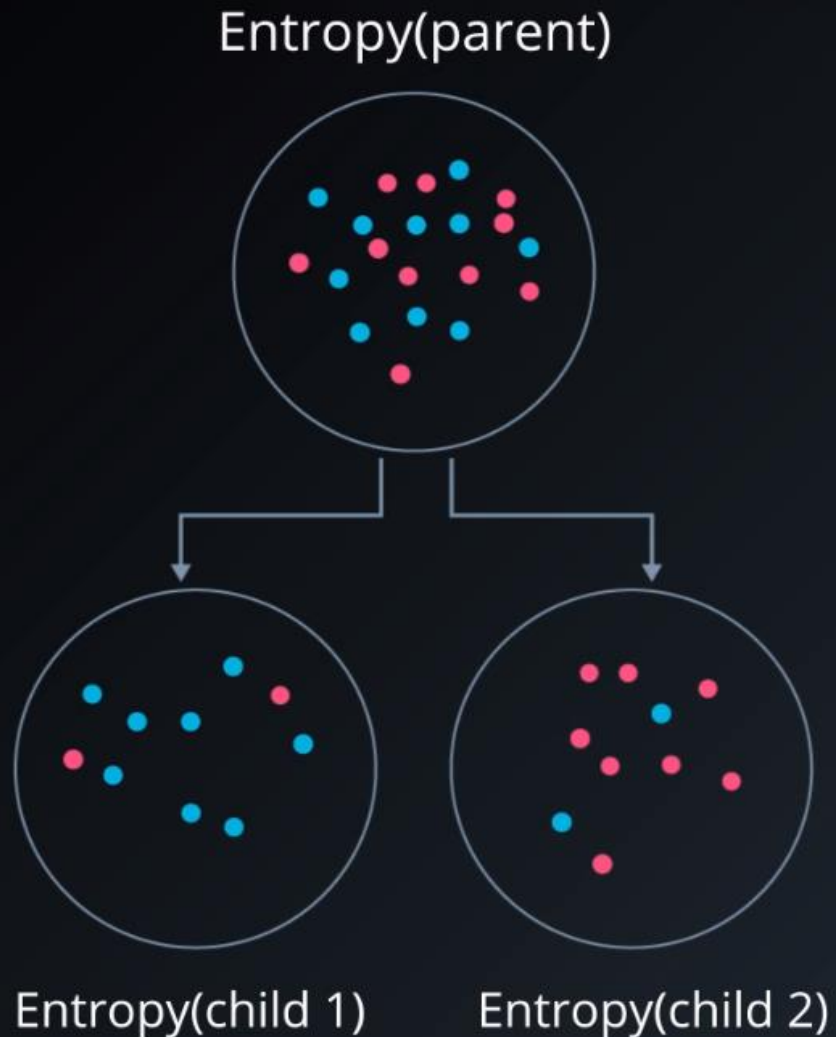
Information gain = 0.28



Information gain = 1



# Ganancia de información (Information Gain)



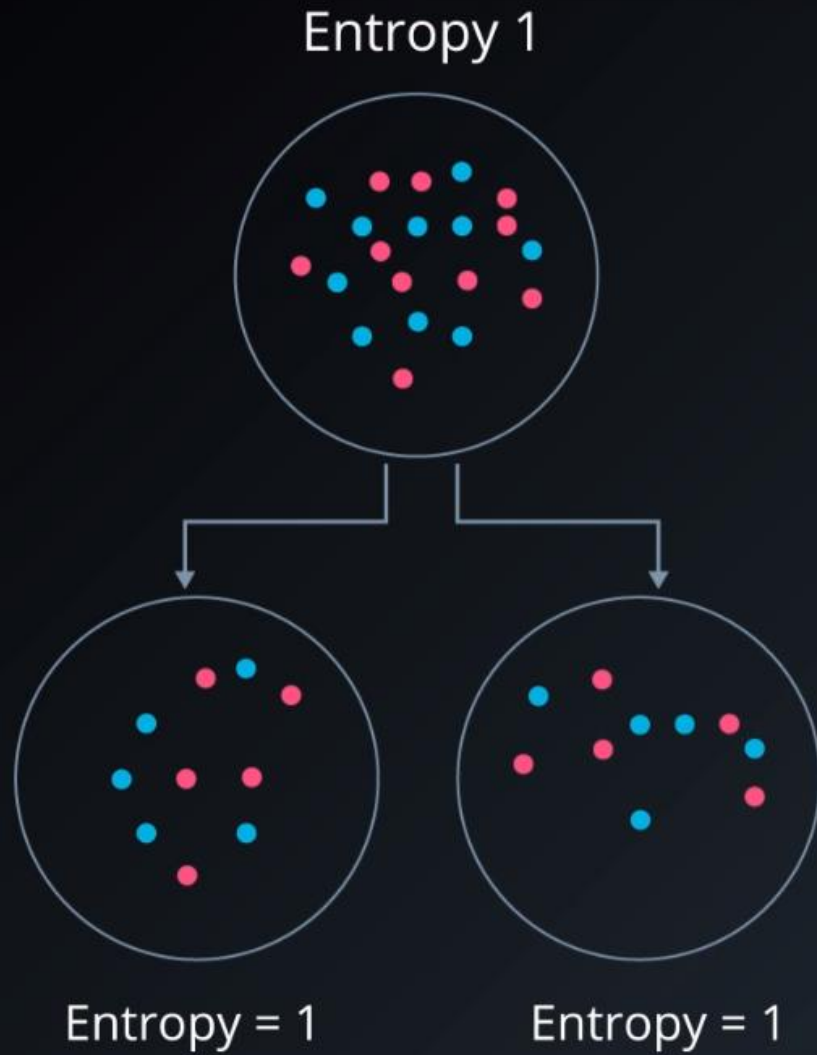
$$\text{Information Gain} = \text{Entropy}(\text{parent}) - 0.5 [\text{Entropy}(\text{child 1}) + \text{Entropy}(\text{child 2})]$$

# Ganancia de información (Information Gain)



Information gain  
 $1 - 0.72 = 0.28$

# Ganancia de información (Information Gain)



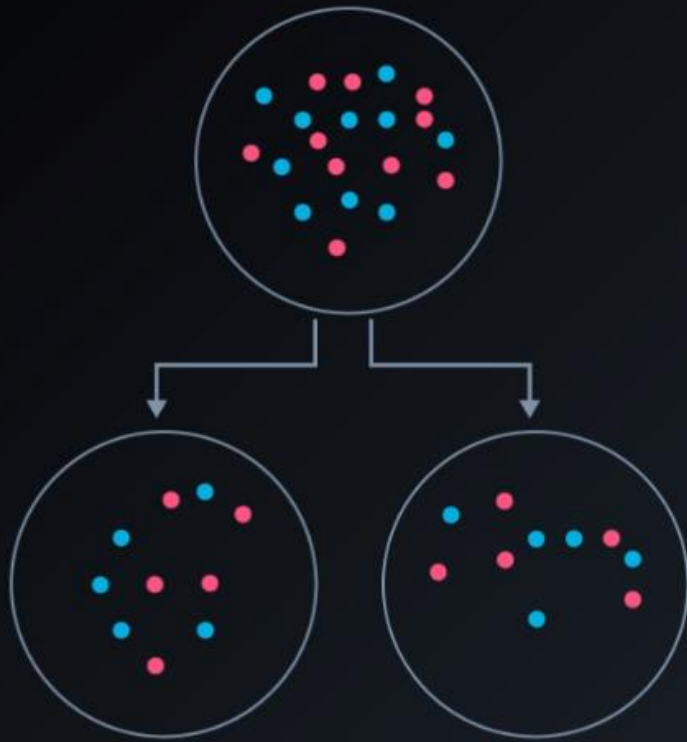
Information gain  
 $1 - 1 = 0$

# Ganancia de información (Information Gain)

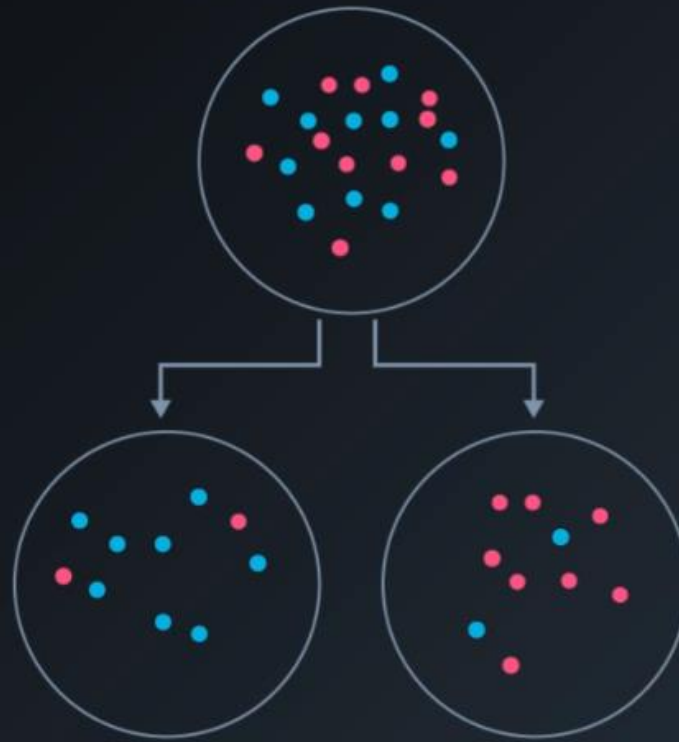


Information gain  
 $1 - 0 = 1$

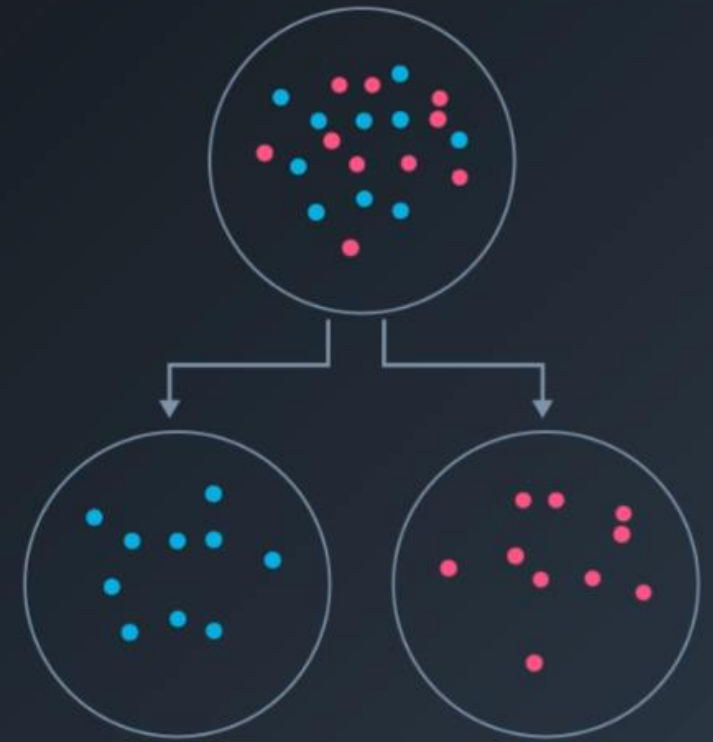
# Ganancia de información (Information Gain)



Information gain = 0












Information gain = 0.28



Information gain = 1

# Maximizando la Ganancia de información







Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

$$\text{Entropy} = -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right) - \frac{1}{6}\log_2\left(\frac{1}{6}\right)$$
$$= 1.46$$

# Maximizando la Ganancia de información

## Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

Gender

F



M



Entropy

0.92





0.92

Information gain =  $1.46 - 0.92 = 0.54$



# Maximizando la Ganancia de información

## Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

### Occupation

S

M



Entropy







0

0.92

$$\text{Information gain} = 1.46 - 0.46 = 1$$

# Maximizando la Ganancia de información

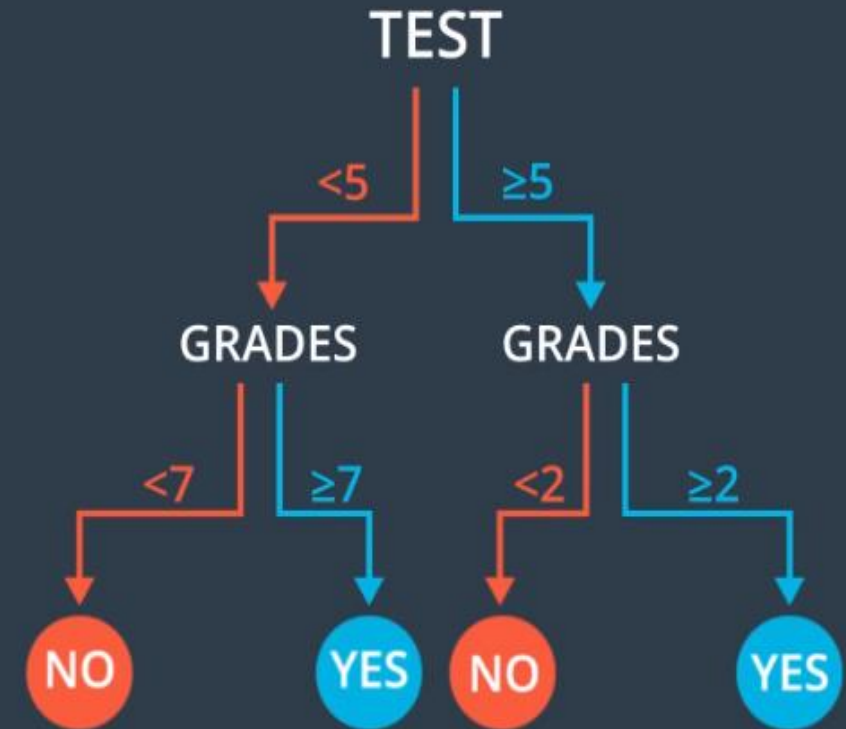
## Recommending Apps

0.54		1
Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

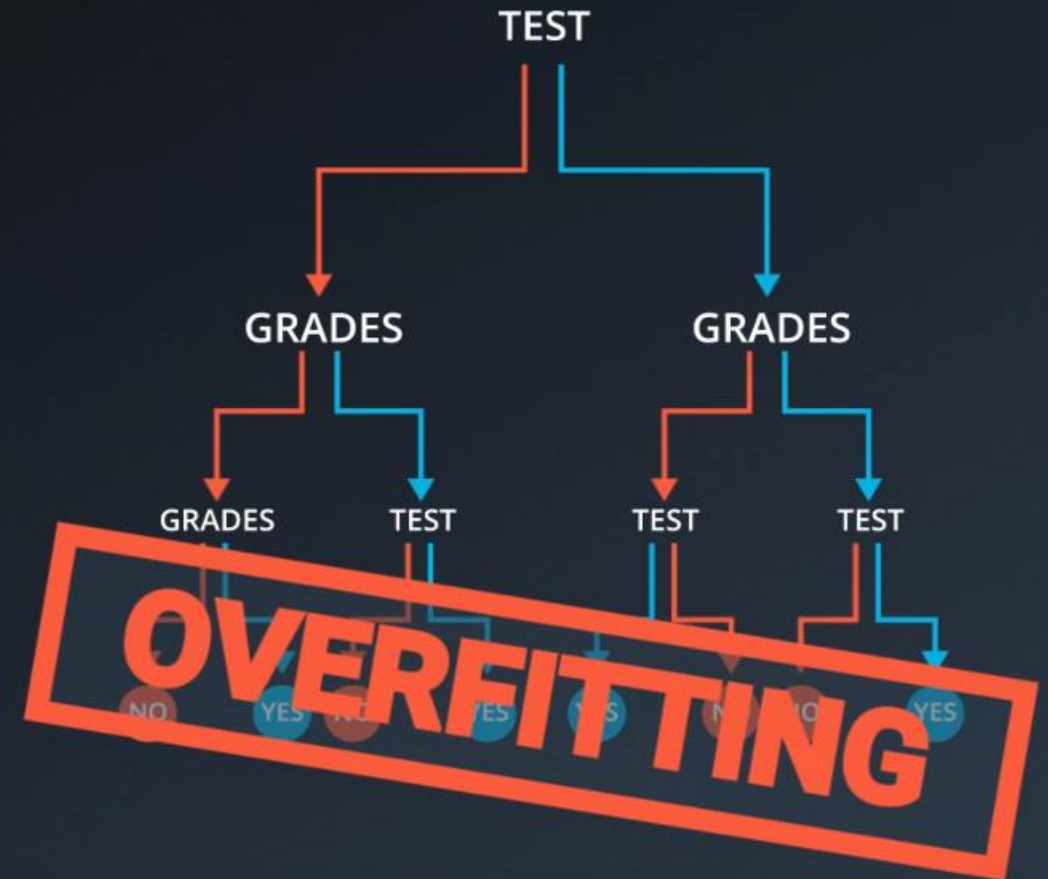
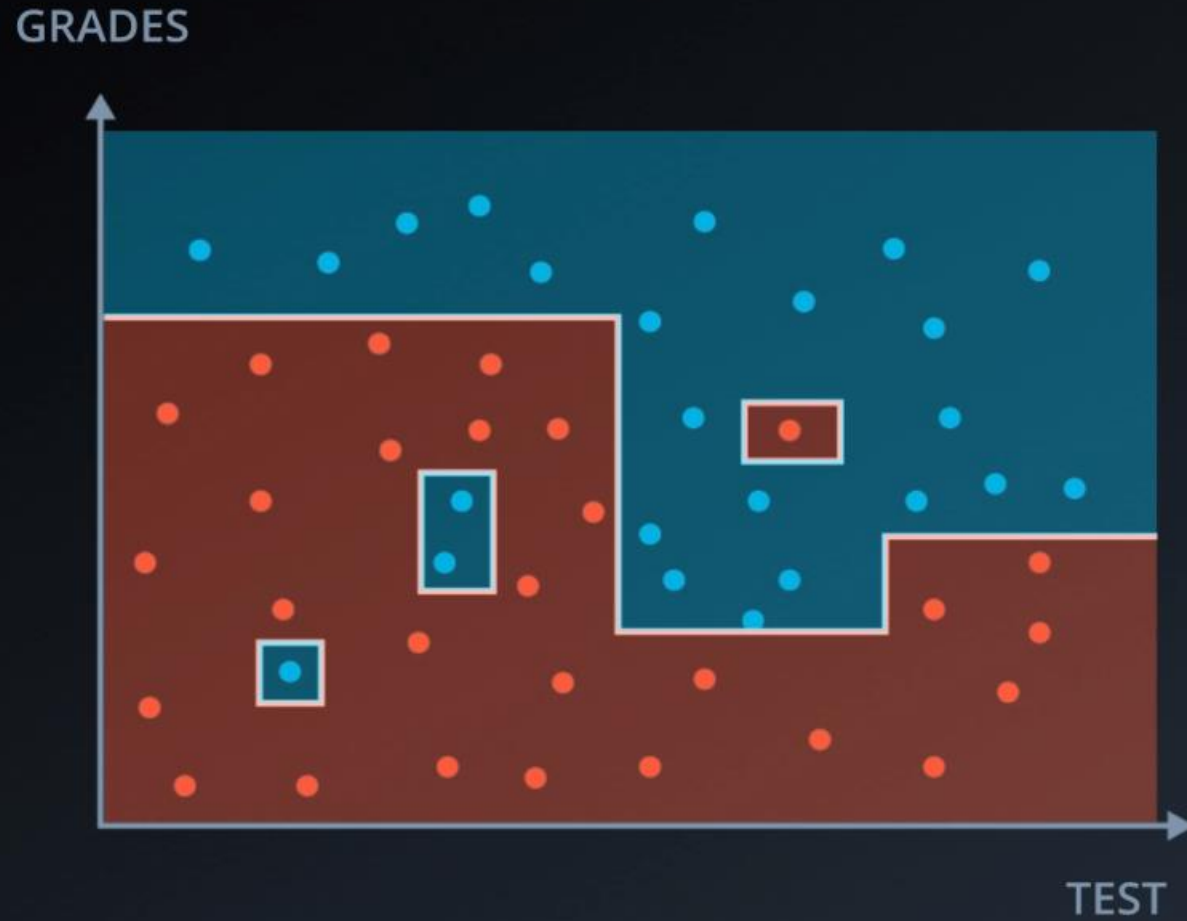
# Maximizando la Ganancia de información



# Maximizando la Ganancia de información



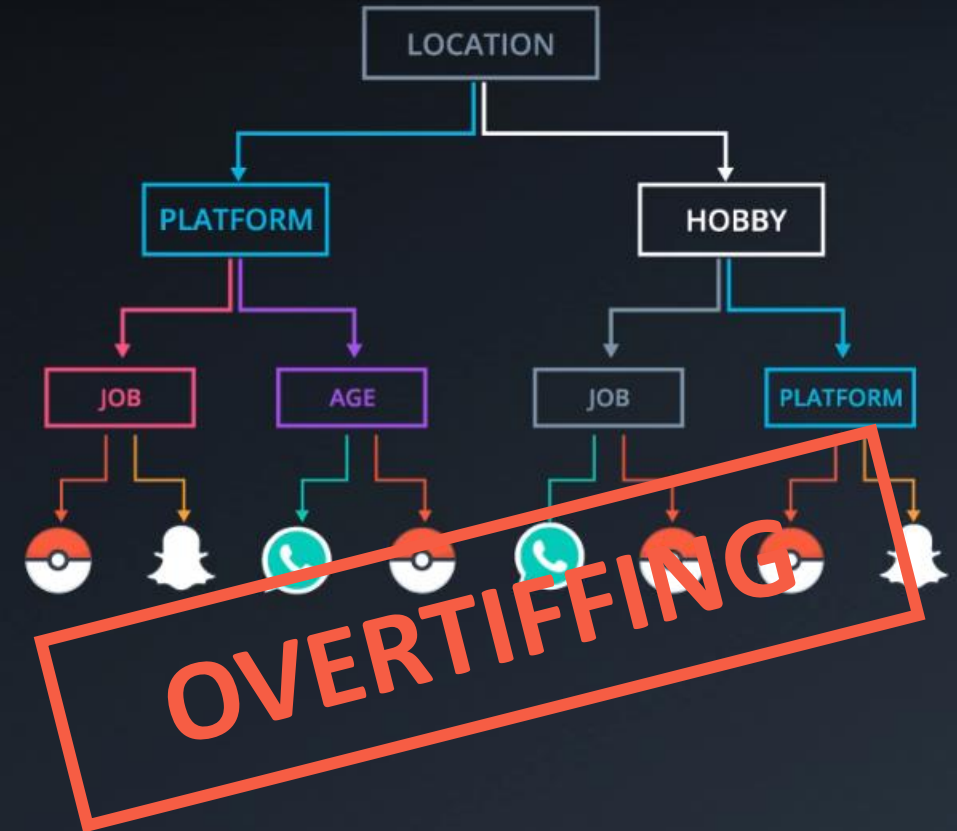
# Overfitting



# Overfitting

## Large Tables

Gender	Age	Location	Platform	Job	Hobby	App
F	15	US	iOS	School	Videogames	
F	25	France	Android	Work	Tennis	
M	32	Chile	iOS	Temp	Tennis	
F	40	China	iOS	Retired	Chess	
M	12	US	Android	School	Tennis	
M	14	Australia	Android	School	Videogames	

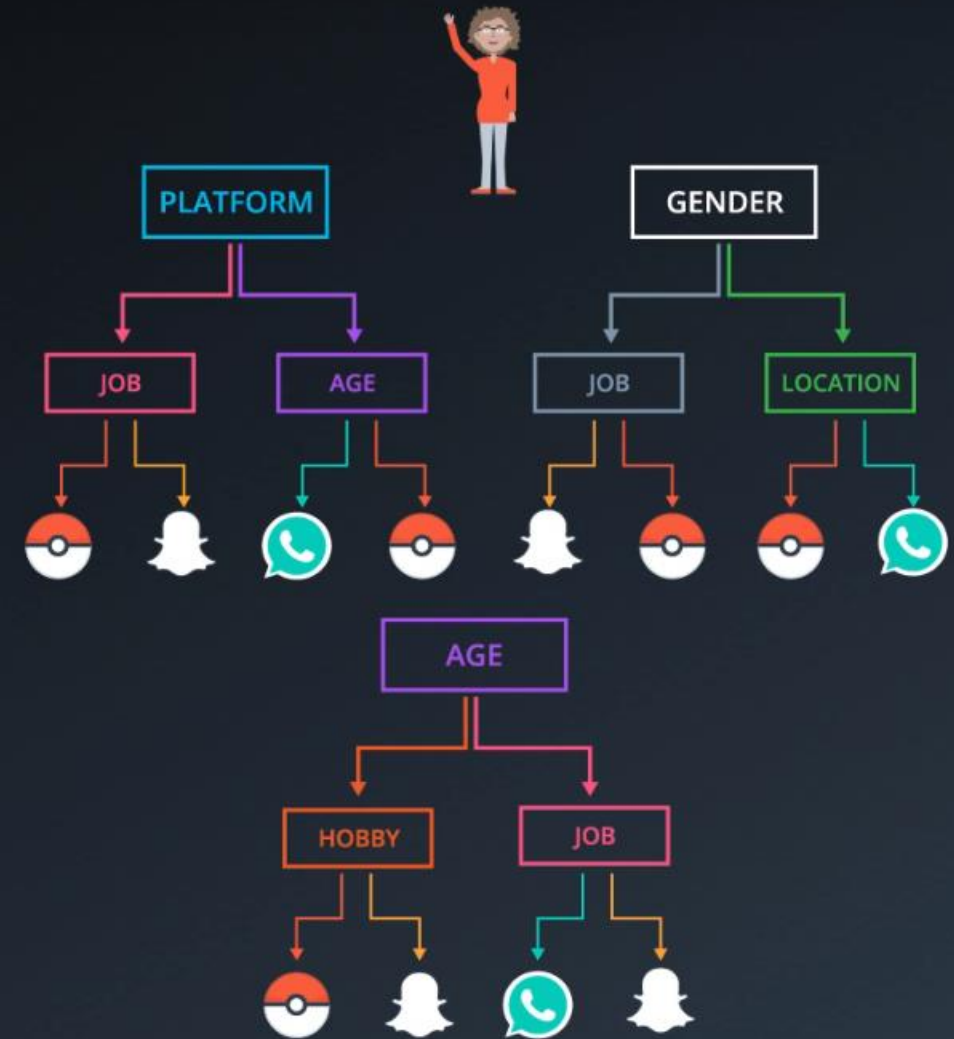




# Random Forests

## Random Forests

Gender	Age	Location	Platform	Job	Hobby	App
F	15	US	iOS	School	Videogames	
F	25	France	Android	Work	Tennis	
M	32	Chile	iOS	Temp	Tennis	
F	40	China	iOS	Retired	Chess	
M	12	US	Android	School	Tennis	
M	14	Australia	Android	School	Videogames	



# Hiperparámetros para Árboles de Decisión

## Maximum Depth (Profundidad máxima)

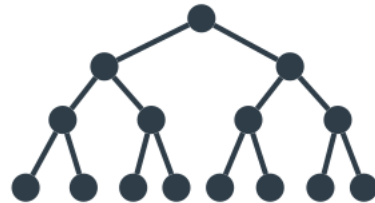
La profundidad máxima de un árbol de decisión es simplemente la longitud más grande entre la raíz y una hoja. Un árbol de longitud máxima  $k$  puede tener como máximo  $2^k$  hojas.



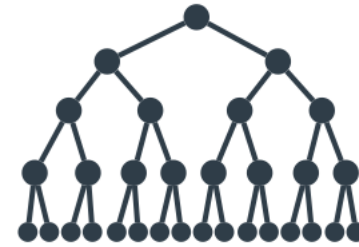
Depth = 1



Depth = 2



Depth = 3



Depth = 4

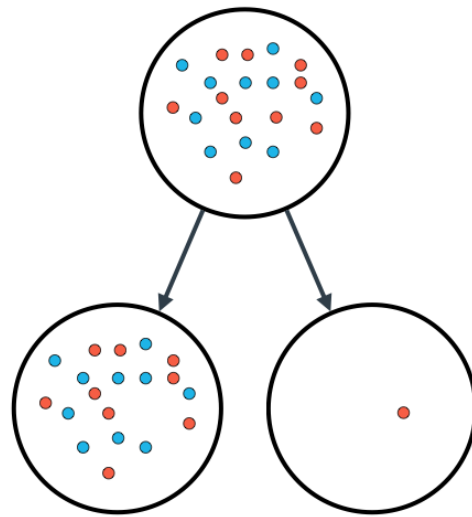


# Hiperparámetros para Árboles de Decisión

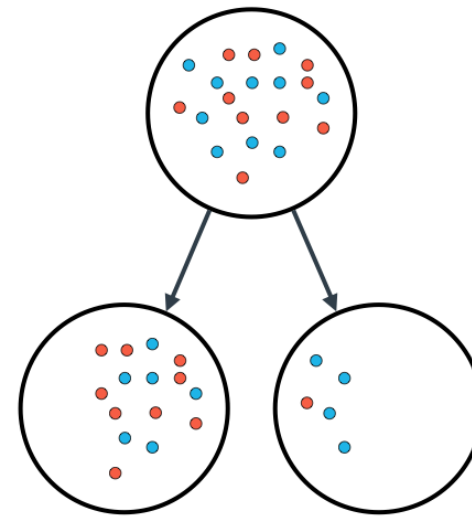
## **Minimum number of samples per leaf** (Número mínimo de muestras por hoja)

Cuando se divide un nodo, uno podría tener el problema de tener 99 muestras en una de ellas y 1 en la otra. Esto no nos llevará demasiado lejos en nuestro proceso, y sería una pérdida de tiempo y recursos.

Si queremos evitar esto, podemos establecer un mínimo para el número de muestras que permitimos en cada hoja.



Minimum samples per leaf = 1



Minimum samples per leaf = 5

# Hiperparametros para Árboles de Decisión

## **Minimum number of samples per split (Número mínimo de muestras por división)**

Este es el mismo que el número mínimo de muestras por hoja, pero se aplica en cualquier división de un nodo.

## **Maximum number of features (Número máximo de características)**

A menudo, tendremos demasiadas características para construir un árbol. Si este es el caso, en cada división, tenemos que verificar el conjunto de datos completo en cada una de las funciones. Esto puede ser muy costoso. Una solución para esto es limitar la cantidad de funciones que uno busca en cada división. Si este número es lo suficientemente grande, es muy probable que encontremos una buena característica entre las que buscamos

# Árboles de decisión en Sklearn



The screenshot shows the scikit-learn website. At the top, there's a navigation bar with links: Home, Installation, Documentation, and Examples. Below this is a grid of 12 small plots showing various machine learning results. To the right of the grid, the text 'scikit-learn' is prominently displayed, followed by 'Machine Learning in Python'. Below this, a list of features is provided:

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

## Classification

Identifying to which category an object belongs to.

**Applications:** Spam detection, Image recognition.

**Algorithms:** SVM, nearest neighbors, random forest, ... — Examples

## Regression

Predicting a continuous-valued attribute associated with an object.

**Applications:** Drug response, Stock prices.

**Algorithms:** SVR, ridge regression, Lasso, ... — Examples

## Clustering

Automatic grouping of similar objects into sets.

**Applications:** Customer segmentation, Grouping experiment outcomes

**Algorithms:** k-Means, spectral clustering, mean-shift, ... — Examples

## Dimensionality reduction

Reducing the number of random variables to consider.

**Applications:** Visualization, Increased efficiency

**Algorithms:** PCA, feature selection, non-negative matrix factorization. — Examples

## Model selection

Comparing, validating and choosing parameters and models.

**Goal:** Improved accuracy via parameter tuning

**Modules:** grid search, cross validation, metrics. — Examples

## Preprocessing

Feature extraction and normalization.

**Application:** Transforming input data such as text for use with machine learning algorithms.

**Modules:** preprocessing, feature extraction. — Examples



# Árboles de decisión en Weka

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

20:45:48 - trees.J48

Classifier output

Number of Leaves : 5

Size of the tree : 9

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	144	96	%
Incorrectly Classified Instances	6	4	%
Kappa statistic	0.94		
Mean absolute error	0.035		
Root mean squared error	0.1586		
Relative absolute error	7.8705 %		
Root relative squared error	33.6353 %		
Total Number of Instances	150		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	M
	0,980	0,000	1,000	0,980	0,990	0
	0,940	0,030	0,940	0,940	0,940	0
	0,960	0,030	0,941	0,960	0,950	0
Weighted Avg.	0,960	0,020	0,960	0,960	0,960	0

=== Confusion Matrix ===

a	b	c	<-- classified as
49	1	0	a = Iris-setosa
0	47	3	b = Iris-versicolor
0	2	48	c = Iris-virginica

Weka Classifier Tree Visualizer: 20:45:48 - trees.J48 (iris)

Tree View

```
graph TD
    Root(petalwidth) -- "<= 0.6" --> L1[Iris-setosa (50.0)]
    Root -- "> 0.6" --> N1(petalwidth)
    N1 -- "<= 1.7" --> N2(petallength)
    N1 -- "> 1.7" --> L2[Iris-virginica (46.0/1.0)]
    N2 -- "<= 4.9" --> L3[Iris-versicolor (48.0/1.0)]
    N2 -- "> 4.9" --> N3(petalwidth)
    N3 -- "<= 1.5" --> L4[Iris-virginica (3.0)]
    N3 -- "> 1.5" --> L5[Iris-versicolor (3.0/1.0)]
```

# Árboles de decisión en Orange

