Classification Problems Sub1

Iván G Torre

Iván G Torre: ivan.gonzalez.torre@upm.es

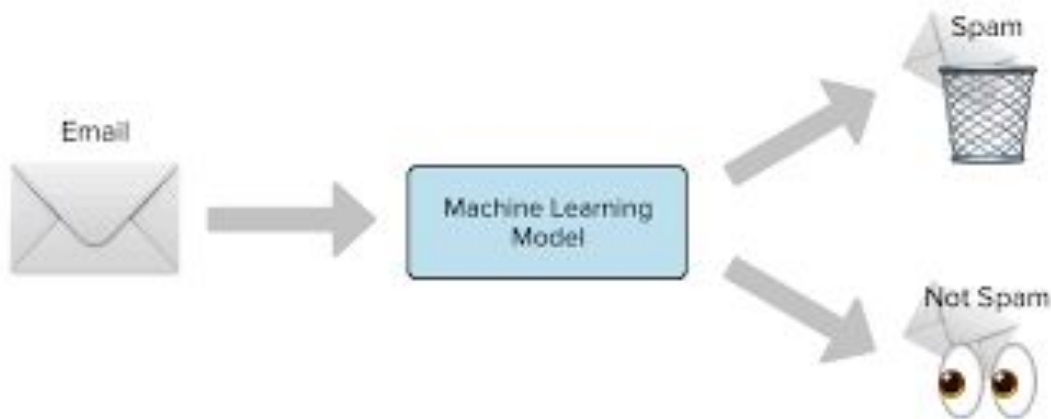Jorge Calero: jorge.calero.sanz@alumnos.upm.es

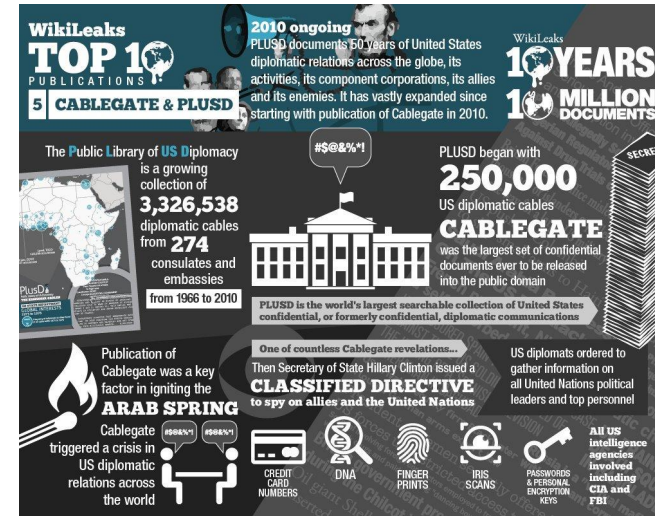José Olarrea: jose.olarrea@upm.es

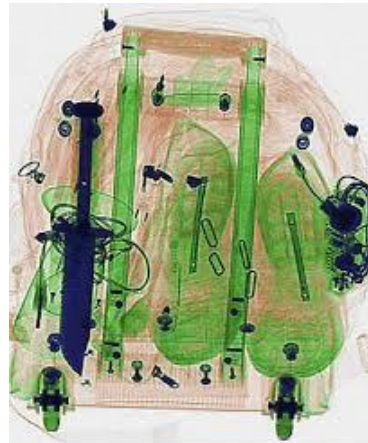# Problemas de clasificación: Ejemplos

Problemas de clasificación: Ejemplos

- Detector de spam (T/F)

Problemas de clasificación: Ejemplos
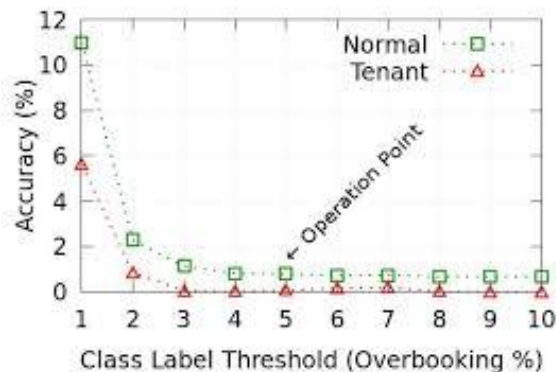
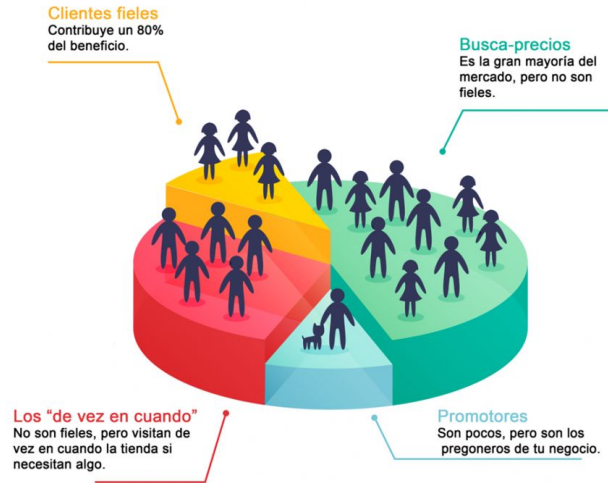- Detector de amenazas (T/F)

Problemas de clasificación: Ejemplos

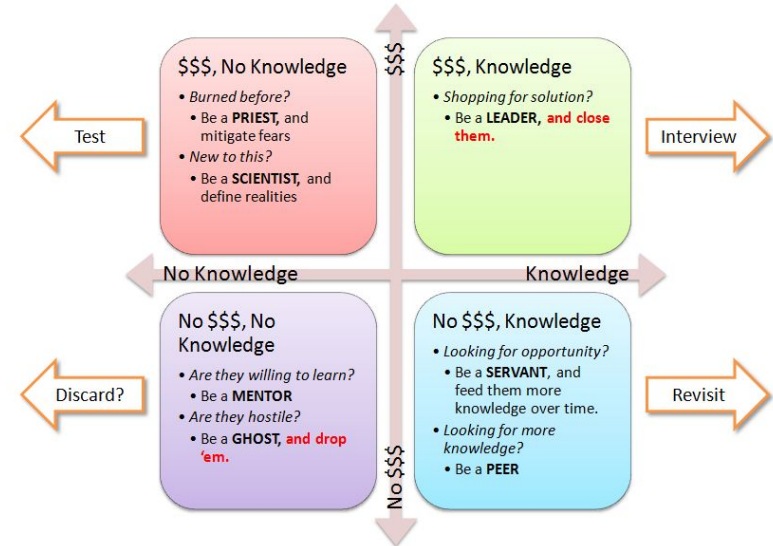- Detector de amenazas (T/F)
  - https://www.youtube.com/watch?v=pNf4-d6fDoY

Problemas de clasificación: Ejemplos

- Overbooking

Name	Description	Type	Number of Values	Information Gain
No-show label	Target label (Y == no-show)	nominal	2	NA
Passenger_Ticketed	Ticket number issued? (Y/N)	nominal	2	0.0563
Frequent_Flier	Air Canada Frequent Flier? (Y/N)	nominal	2	0.0162
Arrival_Airport	Identifies unique flight leg	nominal	79	0.0112
Flight_Leg_Group	Flight-legs grouped by no-show rate	nominal	5	0.0092
PNR_Originator	Booking originator (e.g. AA)	nominal	5	0.0087
Hist_Noshow_Rate	Cabin no-show rate from historical model	real	(5 bins)	0.0079
Booking_Class	Booking class	nominal	11	0.0044
Booking_Class_By_Noshow	Booking class by no-show rate	nominal	3	0.0040
Segment_distance	Segment distance	real	(5 bins)	0.0034
Gender	Derived from title (including unknown)	nominal	3	0.0031
Booking_Class_By_Yield	Booking class by yield	nominal	3	0.0018
Binned_Days_Since_PNR_Creation	Binned days since PNR was created	nominal	5	0.0014
Binned_Departure_Hour	Binned flight departure hour	nominal	4	0.0012
Connecting_Flight	Connecting flight? (Y/N)	nominal	2	0.0007
Binned_Num_Passengers	Number of passengers in this PNR	nominal	3	0.0006
Special_Meal	Ordered special-meal? (Y/N)	nominal	2	0.0005
Binned_Churn	Binned churn values	nominal	3	0.0004
Binned_Num_Segments	Number of flight segments in this PNR	nominal	3	0.0003
Departure_Day	Departure day of week	nominal	7	0.0002

Problemas de clasificación: Ejemplos
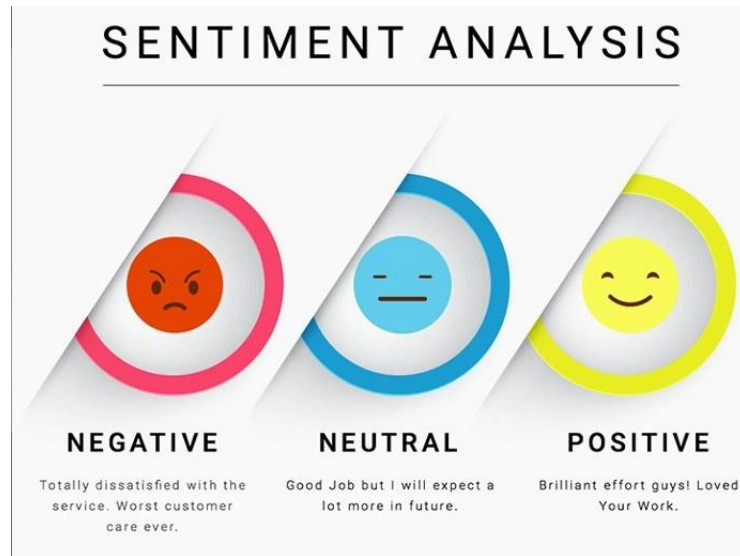
- Segmentación de clientes (Multiclase)

Clientes fieles
Contribuye un 80% del beneficio.

Busca-precios
Es la gran mayoría del mercado, pero no son fieles.

Los "de vez en cuando"
No son fieles, pero visitan de vez en cuando la tienda si necesitan algo.

Promotores
Son pocos, pero son los pregoneros de tu negocio.

Client Segmentation Quadrant

$$$, No Knowledge
- *Burned before?*
  - Be a **PRIEST,** and mitigate fears
- *New to this?*
  - Be a **SCIENTIST,** and define realities

$$$, Knowledge
- *Shopping for solution?*
  - Be a **LEADER, and close them.**

No $$$, No Knowledge
- *Are they willing to learn?*
  - Be a **MENTOR**
- *Are they hostile?*
  - Be a **GHOST, and drop 'em.**

No $$$, Knowledge
- *Looking for opportunity?*
  - Be a **SERVANT,** and feed them more knowledge over time.
- *Looking for more knowledge?*
  - Be a **PEER**

No Knowledge — Knowledge

$$$ — No $$$

Test · Interview · Discard? · Revisit

Problemas de clasificación: Ejemplos

- Voice detection (Multi Clase, (T/F))

"Ok Google"

SENTIMENT ANALYSIS

NEGATIVE
Totally dissatisfied with the service. Worst customer care ever.

NEUTRAL
Good Job but I will expect a lot more in future.

POSITIVE
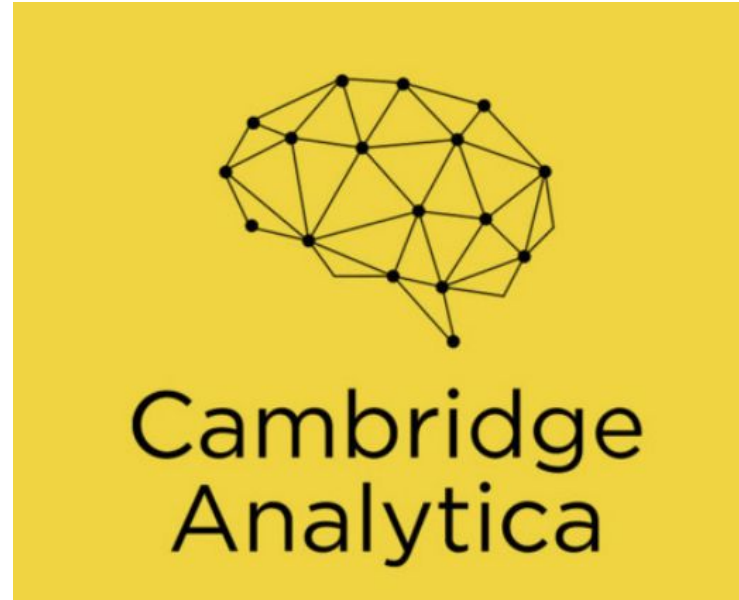Brilliant effort guys! Loved Your Work.

Problemas de clasificación: Ejemplos

- Coche autónomo
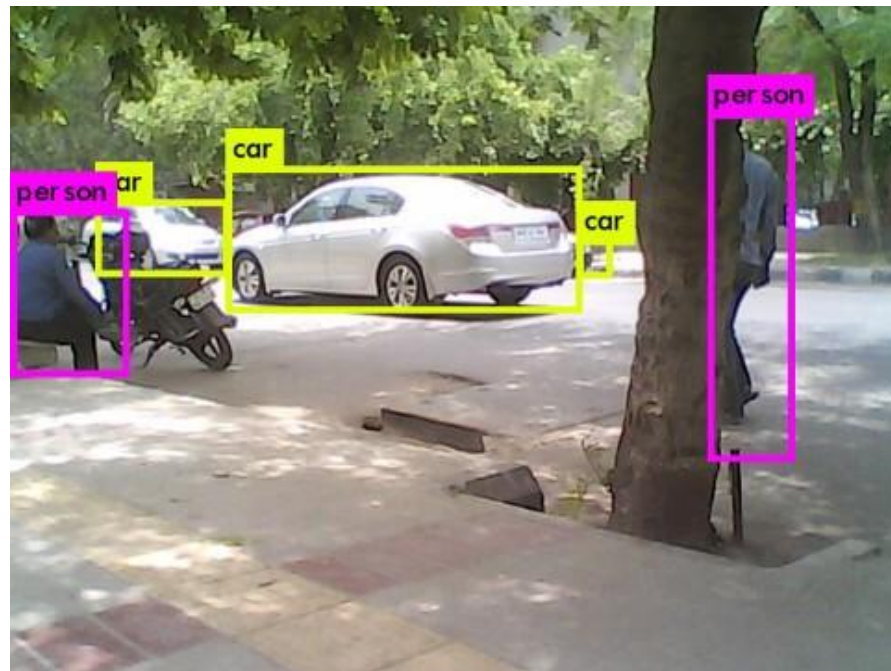
Problemas de clasificación: Ejemplos

- Political segmentation

Problemas de clasificación: Ejemplos

Problemas de clasificación: SPAM

Do You want to score like a young with your wife every time? Know from experience enlarged love productivity EVERY time.

View website

If the button doesn't work, please visit the site:
http://www.osproduction.pl/loggers/webshop/

FedEx®
Federal Express

Dear Valued Customer,

We have been expecting you to contact us for your long existing Package that was registered with us for shipment to your residential location. We thought that the UK National lottery board gave you our contact details to contact us, it may interest you to note that a letter was also added to your FedEx package, however we cannot read the full content to you via email for privacy/security reasons we understand that the package itself is a winning bank Cashier cheque which worth over $1,500.000.00 USD {ONE MILLION FIVE HUNDRED THOUSAND UNITED STATE DOLLAR} As you may know the FedEx Delivery Express Service world-wide do not ship money in CASH but Bank Drafts/winning cheques are shippable. The package is registered with us for mailing by the UK National Lottery Promotion Board, We are sending you this email because your package is been registered on a Special delivery Order on our desk by (U.K NLB). What you have to do now is to contact our Delivery Department for immediate dispatch of your package to your residential address in your country.

Want to score like a pro with your wife every single time? Perceive enlarged passionate output EVERY time.

View website

If You have problems, You can visit the link:
http://www.oneacorn.com/qmdzz3s/webpage/

Attention,

I represent an investor seeking` to invest in any lucrative investment in your country, If you have a solid background and idea of making good profit in real estate or in any business, Please write me for possible business co-operation.

Regards,

Samuel Ayisi

3 Whitehall Court London
SW1A 2EL United Kingdom
Attention:Beneficairy.

NOTE: If you received this message in your SPAM/BULK folder, it is because of the restrictions imposed by your Mail/Internet Service Provider, we urge you to treat it genuinely.

How are you today? Hope all is well with you and family? You may not understand why this mail came to you. I wish to inform you that we have been having a meeting with the IMF for the past 1 month now which ends yesterday in regards to innocent individual who has been a victim of scam that was going on in African and Europe, this Organization(UNITED NATIONS) have agreed to compensate them with the sum of 750,000.00.

This includes every foreign contractors that may have not received their Contract sum, and some people that have an unfinished Transaction/International Businesses that failed due to one problem or the other.

Going through our data base your email contacts and name was found on the list and that is why we are contacting you for you to receive your own compensation and this said money will be paid via ATM MASTER CARD Payment Method which is the easiest and safer way of receiving without any hindrance and it will be issued to you from our Authorized Bank in Nigeria(Africa) You are advised to contact

Mr.Herbert Wigwe
Email: mr.herbertwigwe1201@gmail.com

As he is our Legal Representative in Nigeria, contact him immediately for your own ATM MASTER CARD, you should send him your full Name and telephone number your confidential address where you want him to send the Card.

Hoping to hear from you as soon as you receive your MASTER CARD.

Best regards,
Secretary-General
Antynio Guterres

Cosas raras en un mail

Como comienza todo...

- Statistical methods,
- Bayes

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$$
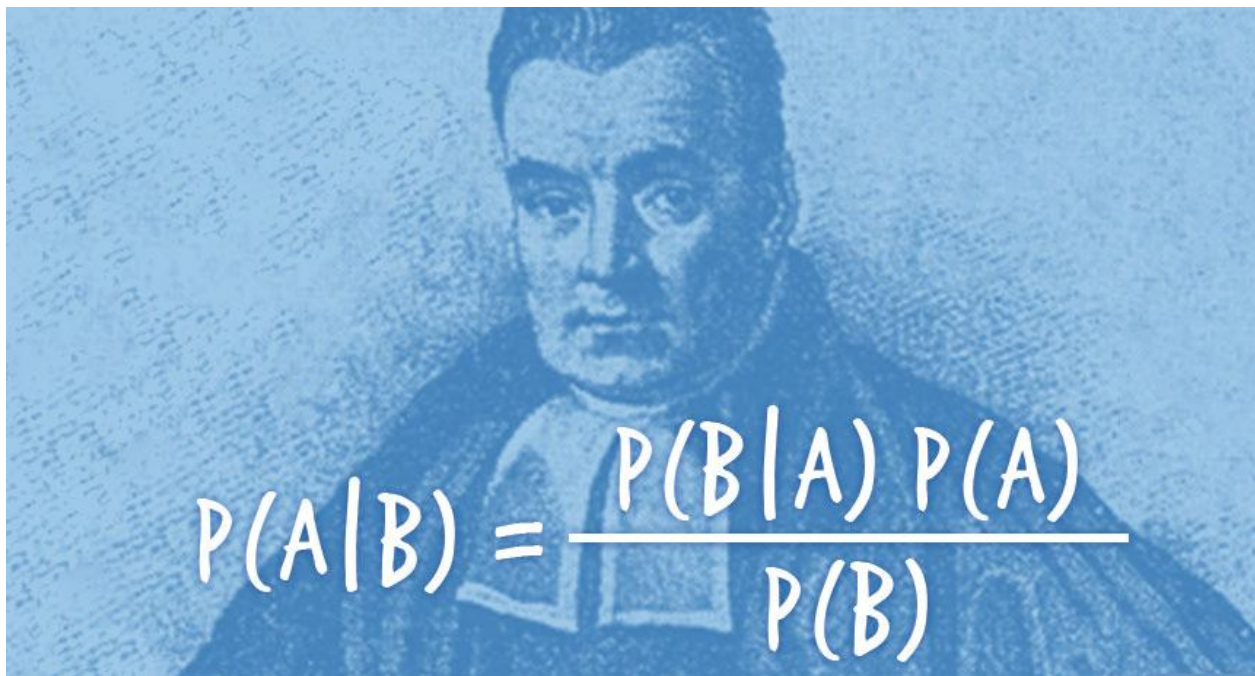
Bayes

- ¿Probabilidad de sacar un as en una baraja?

Bayes

- ¿Probabilidad de sacar un as en una baraja?
  - 40 cartas en la baraja
  - 4 Ases
  - La probabilidad es 4/40

Bayes

- ¿Probabilidad de sacar una figura en una baraja?

Bayes

- ¿Probabilidad de sacar una figura en una baraja?
  - 40 cartas en la baraja
  - 3 (sota, caballo y rey)
  - 4 palos
  - 3*2= 12 Figuras
  - La probabilidad será 12/40

Bayes

- ¿Probabilidad de que la carta escogida sea una figura sabiendo que es un rey?

Bayes

- ¿Probabilidad de que la carta escogida sea una figura sabiendo que es un rey?
    - La carta escogida es un rey
    - Todos los reyes son figuras
    - La probabilidad será 1

Bayes

- ¿Probabilidad de que la carta escogida sea un rey sabiendo que es un figura?

Bayes

- ¿Probabilidad de que la carta escogida sea un rey sabiendo que es un figura?
    - La carta es una figura
    - Conocemos que hay 12 figuras
    - Hay 4 reyes en esas 12 figuras
    - La probabilidad es 4/12

Bayes

● Probabilidad incondicional:

P(As) = 4 Ases /40 cartas

$$P(A) = \frac{No.\ of\ times\ event\ A\ occurs}{Total\ no.\ of\ occurances\ of\ events}$$

Bayes

- Probabilidad Condicional

$$P(B \mid A) = \frac{P(A \mid B) * P(B)}{P(A)}$$

P(Rey|Figura) = P(Figura|Rey) *P(Rey) / P(Figura)

P(Rey|Figura) = 1* (4/40) / (12/40) = 4/12

Detector de spam

La probabilidad de que un mensaje sea Spam si contiene la palabra "herencia" será:

$$\Pr(S|W) = \frac{\Pr(W|S) \cdot \Pr(S)}{\Pr(W|S) \cdot \Pr(S) + \Pr(W|H) \cdot \Pr(H)}$$

La probabilidad de la palabra "herencia" en los mensajes Spam multiplicado por la probabilidad de Spam todo ello dividido por la Probabilidad de que aparezca esa palabra en general.

El denominador explica que la probabilidad de que aparezca la palabra "herencia" es igual a la probabilidad de que aparezca "herencia" en Spam + la prob de que aparezca "herencia" en Ham.

Detector de spam simple: Caso 1

Detector de spam combinando palabras, "bag of words"

the dog is on the table

are	cat	dog	is	now	on	table	the
0	0	1	1	0	1	1	1

$$p = \frac{p_1 p_2 \cdots p_N}{p_1 p_2 \cdots p_N + (1 - p_1)(1 - p_2) \cdots (1 - p_N)}$$

Object → Bag of 'words'

Detector de spam Naive Bayes: Caso 2

https://www.ted.com/talks/james_veitch_this_is_what_happens_when_you_reply_to_spam_email/transcript

Problema de Monty Hall

- Concurso en el que hay 3 puertas: una de ellas tiene premio y las otras dos no. Ahora el concurso transcurre como sigue:
  - El concursante elige una puerta
  - El presentador que conoce lo que hay en las puertas, decide abrir una de las puertas restantes que no tiene premio
  - Pregunta: ¿Quieres conservar tu elección o cambiar?

Problema de Monty Hall
Marilyn vos Savant

"You blew it, and you blew it big! Since you seem to have difficulty grasping the basic principle at work here, I'll explain...

After the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your selection or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don't need the world's highest IQ propagating more. Shame!"

Scott Smith, Ph.D.
University of Florida