



## Sistemas de Ayuda a la Decisión

### Ejercicio sobre Evaluación de Modelos

---

## Índice

1. Objetivos y Descripción de los contenidos	1
2. Materiales disponibles	2
3. Tareas	2
4. Gráficos y fórmulas	3
5. Matrices de Confusión y vectores BOW y tf-idf	5
6. Anexo: Los datos	8

---

## 1. Objetivos y Descripción de los contenidos

Este ejercicio, pretende ser un complemento a la clase teórica, y tiene como objetivo afianzar los conceptos adquiridos en clase sobre las métricas de evaluación de la bonanza de un modelo.

Las **competencias** que el alumno deberá haber adquirido tras realizar este ejercicio son:

1. Capacidad para identificar True Positives (TP), False Positives (FP), True Negatives (TN) y False Negatives (FN).
2. Capacidad para a partir de los valores TP, FP, TN, FN, construir la Matriz de Confusión.
3. Capacidad para calcular la probabilidad a-priori de las clases (o distribución subyacente).
4. Capacidad para calcular las métricas básicas: Accuracy, Precisión, Recall (Sensitivity) y Specificity ( % de true negatives:  $TN \div (TN + FP)$ ) y sus variantes en tareas de clasificación multiclase.
5. Capacidad para obtener un vector de Bag of Words (BOW) o un tf-idf.

## 2. Materiales disponibles

El siguiente material será empleado para el desarrollo de la tarea que se propone y que tiene como objetivo alcanzar las competencias enumeradas anteriormente. Encontrarás los datos al final del documento, en el Anexo.

1. Conjunto de datos de test: Una tabla con 100 instancias de mensajes que han sido clasificados como spam o no spam, donde el 1 indica spam y el 0 no-spam. Las instancias del 20 al 95 son realmente ham y han sido predichas como ham (es decir pares 0 0).
2. Conjunto de datos de test: Una tabla con 20 instancias de mensajes que han sido clasificados como spam o no spam, donde el 1 indica spam y el 0 no-spam.
3. Conjunto de datos de test: Una tabla con 20 instancias de opiniones (reviews) sobre un producto que han sido clasificados como negativas (1), neutras (2) o positivas (3).
4. Gráficos explicativos y fórmulas necesarias

## 3. Tareas

1. Tarea 1: Sobre el conjunto de datos 1
  - a) Sobre la tabla de la tabla 1 marca con color rojo las instancias que son falsos positivos (FP), con verde los verdaderos positivos (TP), con amarillo los Falsos Negativos (FN) y deja sin marcar los verdaderos negativos (TN)
  - b) Construye la Matriz de Confusión a partir de los valores identificados
  - c) Calcula Accuracy, Precision, Recall (Sensitivity), F-Score y Specificity.
2. Tarea 2: Sobre el conjunto de datos 2
  - a) Obten el Bag of Words (BOW) de la primera instancia.
  - b) Obten el tf-idf de la primera instancia.
3. Tarea 3: Sobre el conjunto de datos 3.
  - a) Sobre la tabla 7 marca por cada clase:
    - 1) Opiniones negativas (1)
      - con color rojo las instancias que son falsos positivos (FP)
      - con naranja los verdaderos positivos (TP)
      - con ocre los Falsos Negativos (FN)
      - con amarillo los verdaderos negativos (TN)

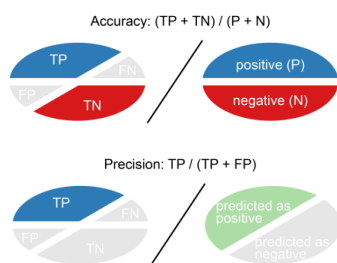
- 2) Opiniones neutras (2)
  - con color morado las instancias que son falsos positivos (FP)
  - con rosa los verdaderos positivos (TP)
  - con lila los Falsos Negativos (FN)
  - con lila claro marcar los verdaderos negativos (TN)
- 3) Opiniones positivas (3)
  - con color azul oscuro las instancias que son falsos positivos (FP)
  - con azul más claro los verdaderos positivos (TP)
  - con azul cielo los Falsos Negativos (FN)
  - con azul casi blanco los verdaderos negativos (TN)
- b) Construye la Matriz de Confusión a partir de los valores identificados
- c) Calcula para cada clase Accuracy, Precision, Recall (Sensitivity), F-Score y Specificity.
- d) Calcula en conjunto macro y micro Precision, macro y micro Recall, macro y micro F-Score.

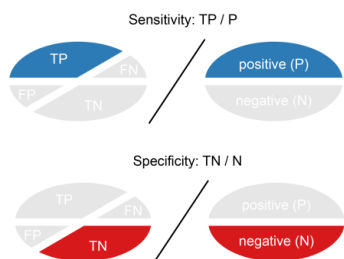
## 4. Gráficos y fórmulas

### Receta:

1. Contar TP (coinciden la predicción positiva y la realidad), FP (la predicción da positivo y la realidad negativo), TN (coinciden realidad y predicción en decir que no es de la clase), FN (la predicción dice que no lo es y la realidad que sí)
2. Obtener la matriz de confusión
3. Calcular las métricas (nota: recall y sensitivity son la misma medida con dos nombres dependiendo del area, ingeniería o medicina)

### Evaluación en clasificación binaria:





$$F-score = \frac{2 * Prec * Recall}{Prec + Recall}$$

### Evaluación en clasificación multiclase:

A parte de la medidas previamente nombradas existen las siguientes:

- Macro-average Precisión: Media de la precisiones de cada clase. Rdo: **Todas clases igual importancia**. Rdo: el impacto de las predominantes se diluye  $\frac{PrecCl1 + PrecCl2 + \dots + PrecCln}{n}$
- Micro-average Precision: Suma la **contribución de todas la instancias** y computa la una media ponderada. Rdo: la precisión de la clase predominante (si en precisión la predicción predominante, si es recall la real predominante) tendrá mayor impacto.  $\frac{TPCl1 + TPCl2 + \dots + TPCln}{(TPCl1 + FPCl1) + (TPCl2 + FPCl2) + \dots + (TPCln + FPCln)}$
- Weighted-average Precision:  $\sum_{i=1}^n Prob(class_i) * Prec_{class_i}$

Primero resolver el siguiente ejemplo

	NEG-Pred	NEUTRO-Pred	POS-Pred	TOTAL REAL
NEG-Real	50	0	0	50
NEUTRO-Real	5	1	0	6
POS-Real	5	0	2	7
TOTAL PRED	60	1	2	

Cuadro 1: Matriz de Confusión de ejemplo

- Precisión:
  - precisiónClase1:  $\frac{50}{60} = 0,83$
  - precisiónClase2:  $\frac{1}{1} = 1$
  - precisiónClase3:  $\frac{2}{2} = 1$
- Recall:

- recallClase1:  $\frac{50}{50} = 1$
- recallClase2:  $\frac{1}{6} = 0,166$
- recallClase3:  $\frac{2}{7} = 0,286$
- F-Score:
  - fScoreClase1:  $\frac{2*precCl1*recallCl1}{precCl1+recallCl1}$
  - fScoreClase2:  $\frac{2*precCl2*recallCl2}{precCl2+recallCl2}$
  - fScoreClase3:  $\frac{2*precCl3*recallCl3}{precCl3+recallCl3}$

## 5. Matrices de Confusión y vectores BOW y tf-idf

	SPAM-Pred	HAM-Pred	TOTAL REAL
SPAM-Real			
HAM-Real			
TOTAL PRED			

Cuadro 2: Matriz de Confusión Tarea 1

	NEG-Pred	NEUTRO-Pred	POS-Pred	TOTAL REAL
NEG-Real				
NEUTRO-Real				
POS-Real				
TOTAL PRED				

Cuadro 3: Matriz de Confusión Tarea 3

0	1	2	3	4	5	6	7
!	,	.	?	congratulation	2	20	2000
8	9	10	11	12	13	14	15
3510i	4	500	accomodate	award	black	busy	call
16	17	18	19	20	21	22	23
cash-balance	chat	class	congratulation	contact	current	dear	dinner
24	25	26	27	28	29	30	31
draw	dream	finish	friday	friend	go	good	guess
32	33	34	35	36	37	38	39
he	hear	home	i	iphone	late	line	logo
40	41	42	43	44	45	46	47
lover	me	motorola	msg	new	nokia	not	now
48	49	50	51	52	53	54	55
number	offer	ok	old	one	owe	phone	pls
56	57	58	59	60	61	62	63
pound	private	prize	remember	rude	send	show	sorry
64	65	66	67	68	69	70	71
thank	today	tomorrow	tonight	try	u	ur	urgent
72	73	74	75	76	77	78	79
video	want	way	we	where	while	win	worry
80	81	82					
xmas	year	you					

Cuadro 4: Vector BOW para la tarea 2.

0	1	2	3	4	5	6	7
!	,	.	?	congratulation	2	20	2000
8	9	10	11	12	13	14	15
3510i	4	500	accomodate	award	black	busy	call
16	17	18	19	20	21	22	23
cash-balance	chat	class	congratulation	contact	current	dear	dinner
24	25	26	27	28	29	30	31
draw	dream	finish	friday	friend	go	good	guess
32	33	34	35	36	37	38	39
he	hear	home	i	iphone	late	line	logo
40	41	42	43	44	45	46	47
lover	me	motorola	msg	new	nokia	not	now
48	49	50	51	52	53	54	55
number	offer	ok	old	one	owe	phone	pls
56	57	58	59	60	61	62	63
pound	private	prize	remember	rude	send	show	sorry
64	65	66	67	68	69	70	71
thank	today	tomorrow	tonight	try	u	ur	urgent
72	73	74	75	76	77	78	79
video	want	way	we	where	while	win	worry
80	81	82					
xmas	year	you					

Cuadro 5: Vector tf-idf para la tarea 2.

## 6. Anexo: Los datos

Figura 1: Muestra de Test (Conjunto de Datos 1)

mensaje	predicción	real/deseado
men1	1	1
men2	1	0
men3	0	0
men4	0	0
men5	1	1
men6	0	0
men7	0	0
men8	1	0
men9	0	0
men10	0	1
men11	1	1
men12	0	0
men13	0	0
men14	1	0
men15	0	0
men16	0	0
men17	0	0
men18	0	0
men19	0	0
men96	0	0
men97	0	0
men98	0	0
men99	0	0
men100	0	0



<b>Id</b>	<b>SMS</b>	<b>Clase Real</b>	<b>Clase Pred.</b>
1	Remember U owe me 20 pounds	0	1
2	Congratulations! One year older! I am calling U...	0	1
3	Sorry, I'll call later	0	0
4	Urgent! Finished class where are you.	0	1
5	K. Did you call me just now ah?	0	0
6	Ok i am on the way to home. Do you offer me dinner tonight?	0	1
7	Dont worry. I guess he's busy.	0	0
8	Do I go home?	0	0
9	Going for dinner. Msg you after.	0	0
10	Dear, will call Tomorrow. Pls accomodate.	0	0
11	Congratulations - Thanks to a good friend U have WON	1	1
12	Send a logo 2 ur lover	1	0
13	Ur cash-balance is currently 500 pounds	1	1
14	URGENT! Your number has been awarded with a prize of 2000 pounds	1	1
15	Do you want a new Video phone?	1	0
16	Urgent! Black friday Offer! Latest Iphone	1	1
17	Not heard from U4 a while. Call 4 rude chat private line	1	0
18	URGENT! We are trying to contact U. Todays draw shows U won 2000 pounds	1	1
19	Do you dream of a New Nokia 3510i	1	0
20	Xmas Offer! Latest Motorola	1	1

Cuadro 6: Ejemplos del Conjunto de Datos 2 (SMS: Clasificación SPAM/HAM) para la tarea de Evaluación 2. Previamente habrá que obtener BOW

<b>Id</b>	<b>atr1</b>	<b>atr2</b>	<b>atr3</b>	<b>atr4</b>	<b>Clase Real</b>	<b>Clase PredMod1</b>	<b>Clase PredMod2.</b>
Opi1	valor	valor	valor	valor	1	1	2
Opi2	valor	valor	valor	valor	1	1	1
Opi3	valor	valor	valor	valor	3	1	2
Opi4	valor	valor	valor	valor	3	1	3
Opi5	valor	valor	valor	valor	2	1	2
Opi6	valor	valor	valor	valor	2	2	2
Opi7	valor	valor	valor	valor	2	2	2
Opi8	valor	valor	valor	valor	3	3	2
Opi9	valor	valor	valor	valor	2	1	2
Opi10	valor	valor	valor	valor	2	1	2
Opi11	valor	valor	valor	valor	3	1	3
Opi12	valor	valor	valor	valor	2	1	2
Opi13	valor	valor	valor	valor	2	1	2
Opi14	valor	valor	valor	valor	3	3	3
Opi15	valor	valor	valor	valor	2	2	2
Opi16	valor	valor	valor	valor	2	1	2
Opi17	valor	valor	valor	valor	2	1	2
Opi18	valor	valor	valor	valor	3	1	3
Opi19	valor	valor	valor	valor	3	3	3
Opi20	valor	valor	valor	valor	3	3	3

Cuadro 7: Ejemplos del Conjunto de Datos 3 (Opiniones: Análisis de sentimientos) para la tarea de Evaluación 3.

## Referencias

- [1] Data Mining: Practical Machine Learning Tools and Techniques Morgan Kaufmann Publishers Inc. (The Morgan Kaufmann Series in Data Management Systems) Ian H. Witten, Eibe Frank, Mark A. Hall
- [2] Takaya Saito and Marc Rehmsmeier (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 10(3):e011843