

# LO QUE INTENTABA SER UN MODELO DE CLASIFICACIÓN PARA LA DETECCIÓN DE PERFILES DE RIESGO RESPECTO AL JUEGO

que luego fueron más, pero no... aunque al final parece que algo sí.

---

**PRESENTADO POR**

David Sanchez Corbella

# HOLA! ME LLAMO DAVID Y SOY CROUPIER

---

- Sí, los de las bolitas y las cartas.
- No, no es un BINGO, es un Casino.
- No, no es tan guay, por eso estoy aquí.
- Para ganar mucho hay que jugar mucho.
- Sí, hay gente que se deja mucho dinero.
- **Y sí, hay gente que está todos los días...  
y de eso trata mi proyecto.**



# TABLA DE CONTENIDOS

---

**4** Objetivo

---

**5** Dataset

---

**6** Limpieza de dataset

---

**7** Creación de la variable target

---

**8** EDA

---

**12** Modelos

---

**21** Conclusiones

---

**23** Contacto

# OBJETIVO

Crear una herramienta a partir de un modelo de clasificación que ayude en la identificación precoz de los problemas de juego de la población.



# DATASET

Estudio de 2015 de la DGOJ sobre prevalencia de juego

6816 rows  
240 columns



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE DERECHOS SOCIALES, CONSUMO  
Y AGENDA 2030

SECRETARÍA GENERAL  
DE CONSUMO Y JUEGO

DIRECCIÓN GENERAL  
DE ORDENACIÓN DEL JUEGO

Dirección General de Ordenación del Juego



Estudios prevalencia / clínico

Año 2022-2023

- Estudio de Prevalencia de Juego 2022-2023
  - Nota metodológica
  - Resumen ejecutivo
  - Cuestionario de la encuesta

Año 2017

- Estudio y análisis de los factores de riesgo del trastorno de juego en población clínica española
- Detalle datos estudio clínico
- Aviso Legal

Año 2015

- Estudio sobre prevalencia, comportamiento y características de los usuarios de juego de azar en España
- Infografía

# LIMPIEZA DE DATASET

	P0B2	P0C	DB0	DB1	DB5	P28	P35	P27	P31	P42	P29	P46	POF3	P3	P4	P5	P8	P20	P21	P22	P23	P24	P1A	I
1	32	1	4	2	2	4	1	2	2	2	3	7	2	1	1	1	1	2	1	2	1	2	2	
2	52	1	3	4	1	3	4	2	2	2	3	8	2	1	1	2	2	2	3	2	1	2	5	
3	67	2	2	4	4	4	1	2	2	2	2	9	2	1	1	2	1	2	3	2	1	2	2	
4	60	2	3	1	2	5	1	2	2	2	2	11	2	1	1	1	2	1	1	2	1	2	5	
5	51	2	3	1	2	4	2	2	2	2	2	9	2	1	1	1	1	1	1	2	1	2	4	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
6811	47	2	3	1	2	4	2	3	2	2	3	9	2	1	1	2	2	1	2	2	1	2	3	
6813	33	2	3	1	1	3	0	2	1	1	1	6	2	1	1	1	1	2	2	2	1	2	5	
6814	39	2	2	6	2	5	0	4	2	1	5	9	2	1	1	1	1	2	2	2	1	2	2	
6815	57	1	3	1	2	4	5	2	2	2	2	9	2	1	1	2	1	1	2	2	1	2	5	
6816	18	1	3	1	1	5	3	3	2	2	3	5	2	1	1	1	1	2	2	2	1	2	2	

4669 rows  
38 columns

POF2 == 1

COLUMNAS

TRANSFORMACIÓN

NULOS

NS/NC

# CREACIÓN DE LA VARIABLE TARGET

3 clasificaciones a partir de 17 preguntas de un estudio clínico:

- Sin riesgo
- Con riesgo
- Con problemas

## ANEXO

```
df_target.loc[(df_target['clasificacion'] == 0) & (df_target['clasificacion2'] == 1), 'clasificacion'] = 1
df_target.loc[(df_target['clasificacion'] == 0) & (df_target['P43'] == 1), 'clasificacion'] = 1
df_target.loc[(df_target['clasificacion'] == 0) & (df_target['P15A'] == 1), 'clasificacion'] = 1
df_target.loc[(df_target['clasificacion'] == 0) & (df_target['P15B'] == 1), 'clasificacion'] = 2
df_target.loc[(df_target['clasificacion'] == 0) & (df_target['P16A'] == 1), 'clasificacion'] = 1
df_target.loc[(df_target['clasificacion'] == 0) & (df_target['P16B'] == 1), 'clasificacion'] = 1

df['clasificacion'] = df_target['clasificacion']
df['clasificacion'].value_counts()
```

```
p14b = df_target.filter(like='P14B')

def calcular_clase(row):
    multiplicador = 0

    for col in p14b.columns:
        if row[col] == 1:
            if col in ['P14B5', 'P14B7', 'P14B12']:
                multiplicador += 9
            elif col in ['P14B3', 'P14B15', 'P14B16']:
                multiplicador += 2
            elif col in ['P14B8', 'P14B9', 'P14B17']:
                multiplicador += 1
            else:
                multiplicador += 3

    if multiplicador < 3:
        return 0
    elif multiplicador < 9:
        return 1
    else:
        return 2

df_target['clasificacion'] = df_target.apply(calcular_clase, axis=1)

df_target['clasificacion'].value_counts()
```

# EDA

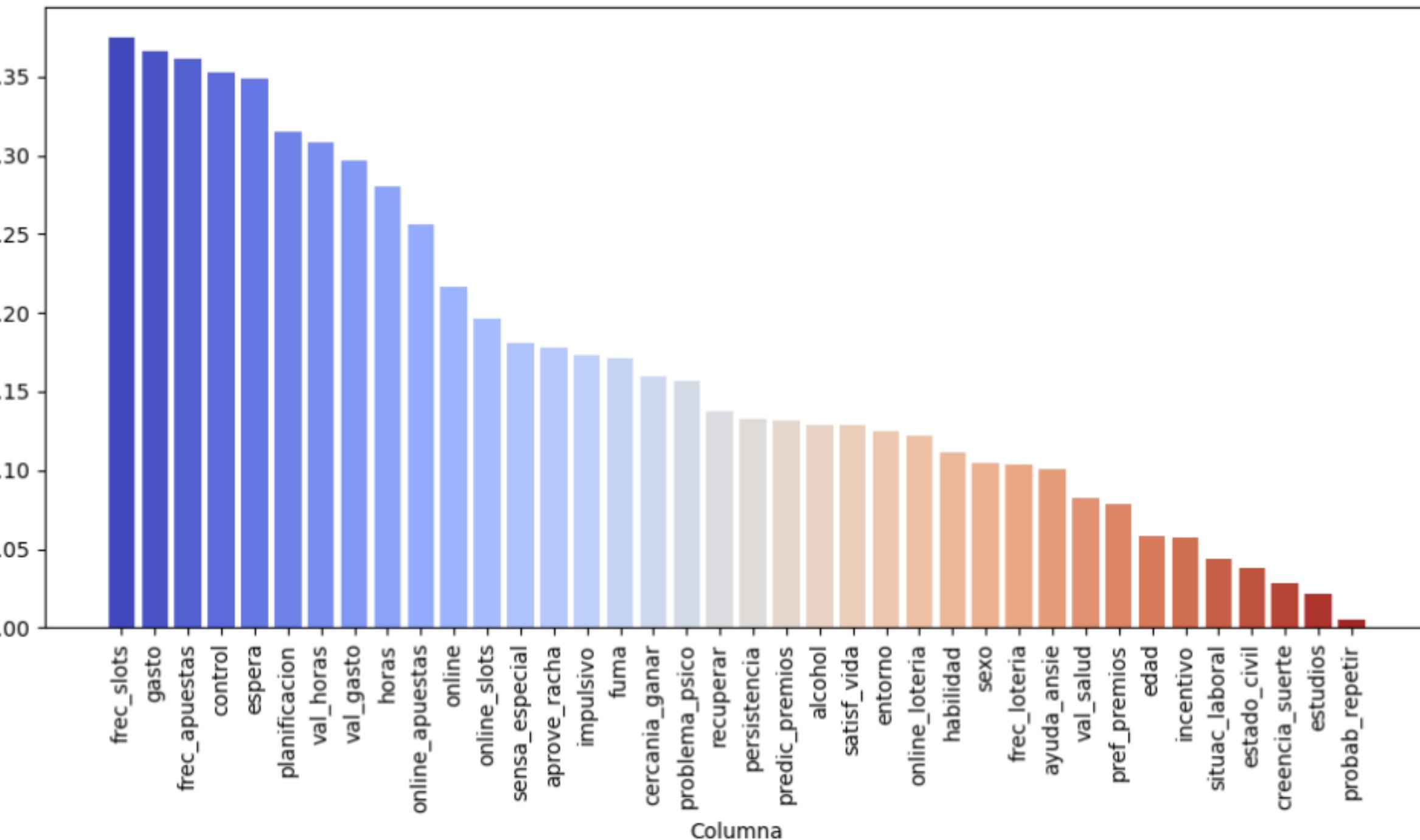
Análisis Exploratorio de Datos

- Correlaciones
- Boxplots
- Kdeplot



# Correlaciones respecto a “clasificación”

frec\_slots  
gasto  
frec\_apuestas  
control  
espera  
planificacion  
val\_horas  
online\_apuestas  
online  
online\_slots



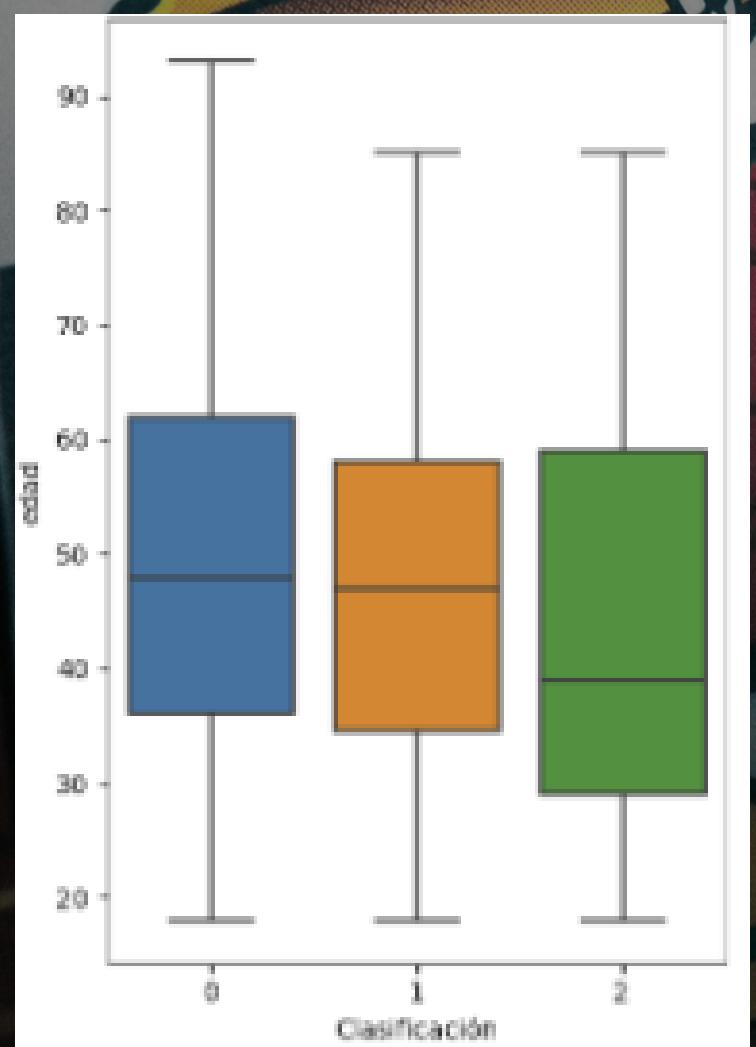
probab\_repetir  
estudios  
creencia\_suerte  
estado\_civil  
situac\_laboral  
incentivo  
edad  
pref\_premios  
val\_salud  
ayuda\_ansie

+15

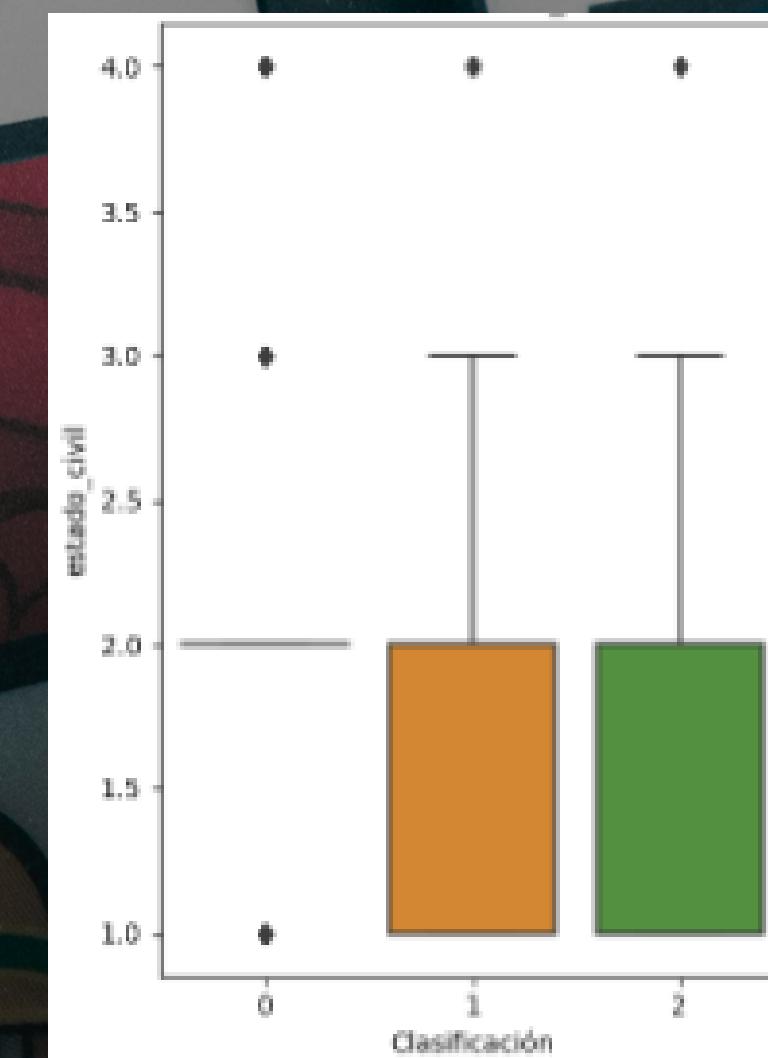
BIG CHANCE は30ゲーム又は3回のBONUS CHANCEで終了

JAC

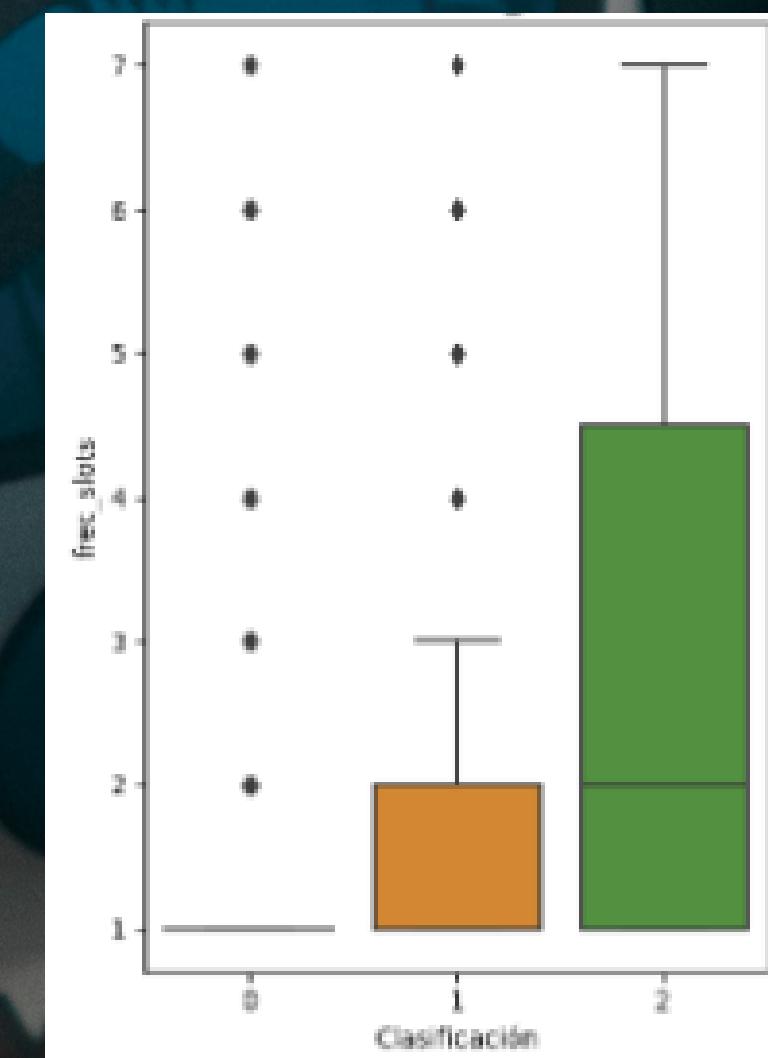
EDAD



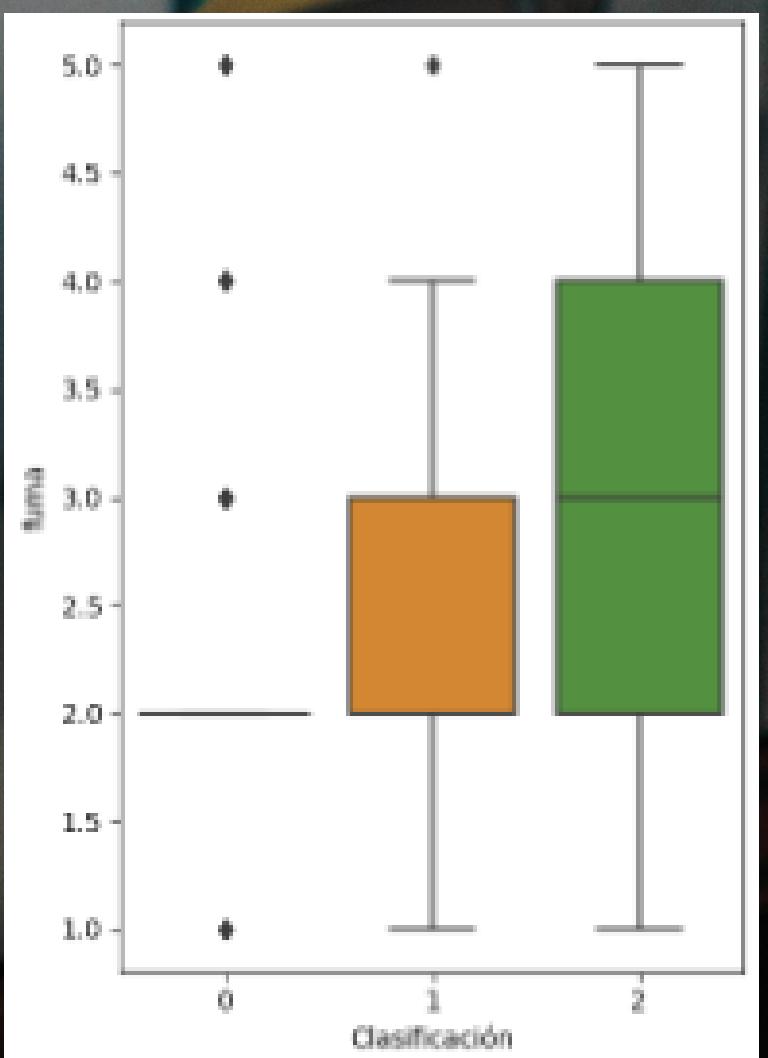
ESTADO\_CIVIL



FREC\_SLOTS

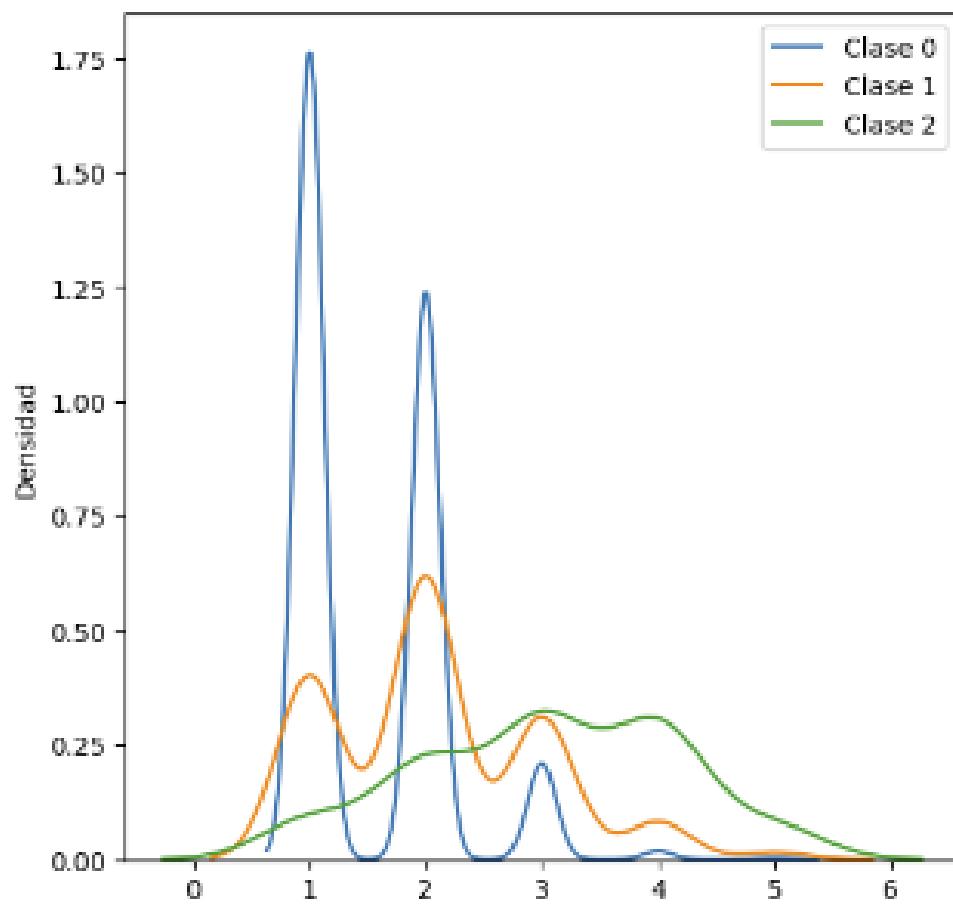


FUMA



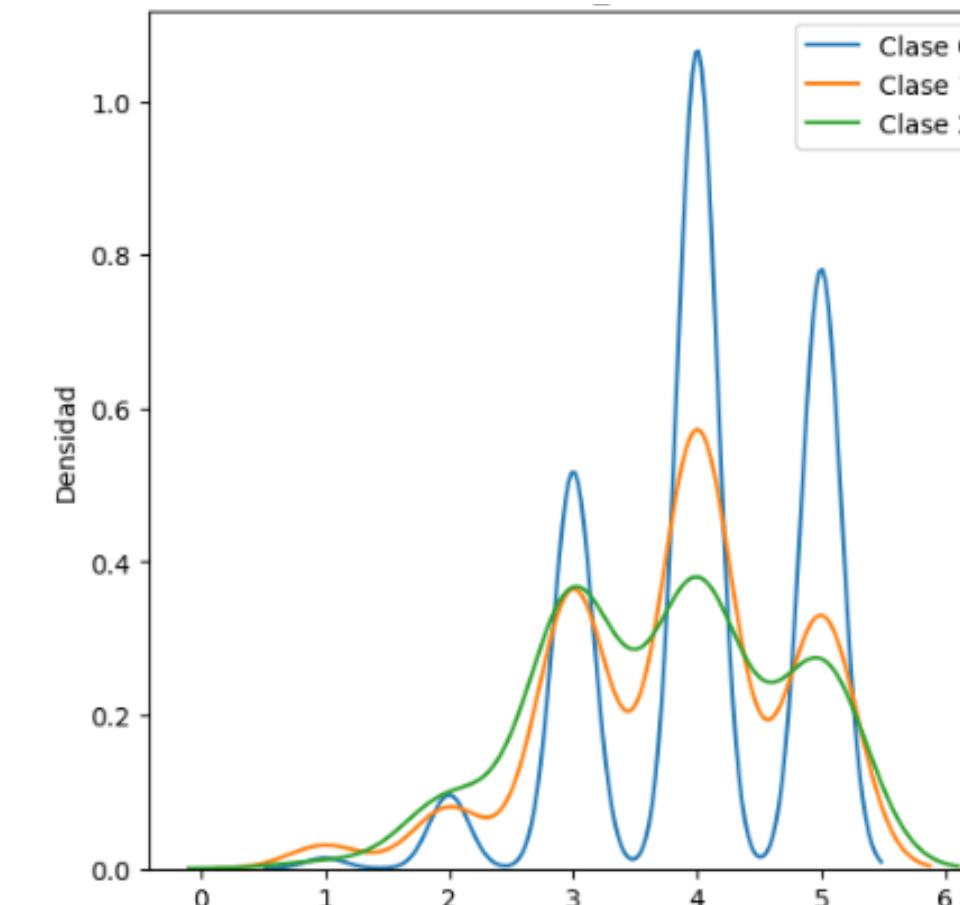
## GASTO

Distribución de gasto por clase



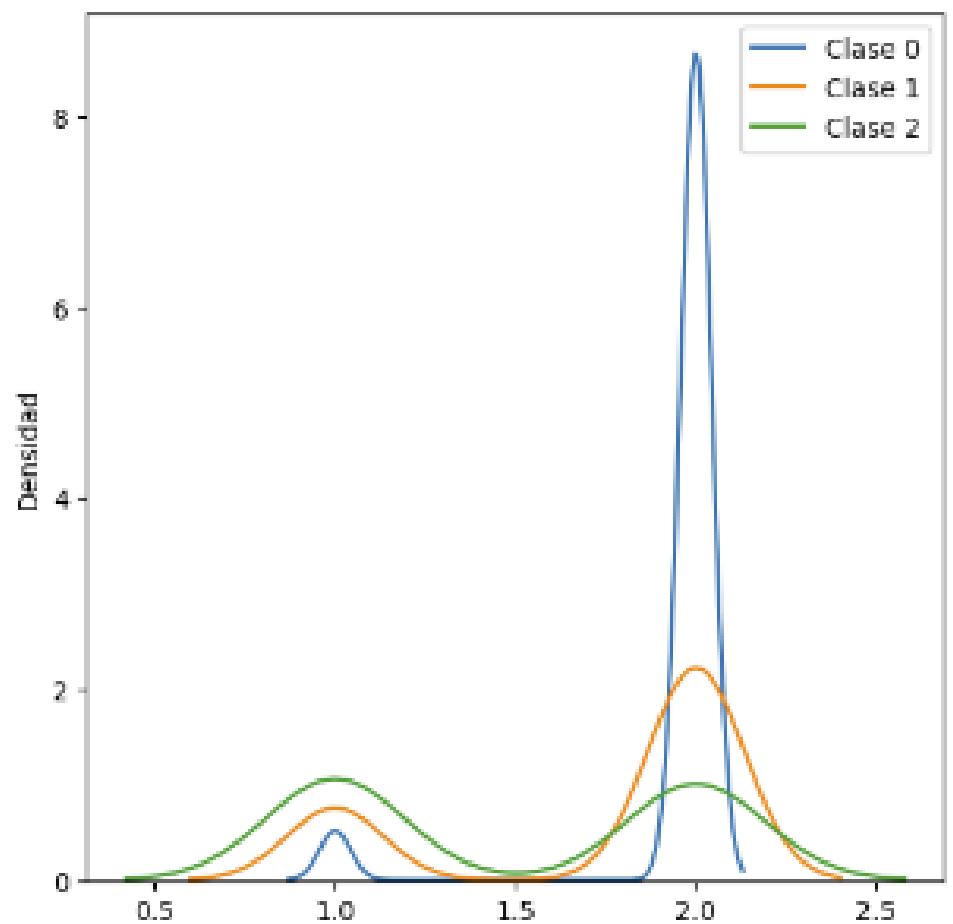
## VAL\_SALUD

Distribución de la valoración de la salud por clase



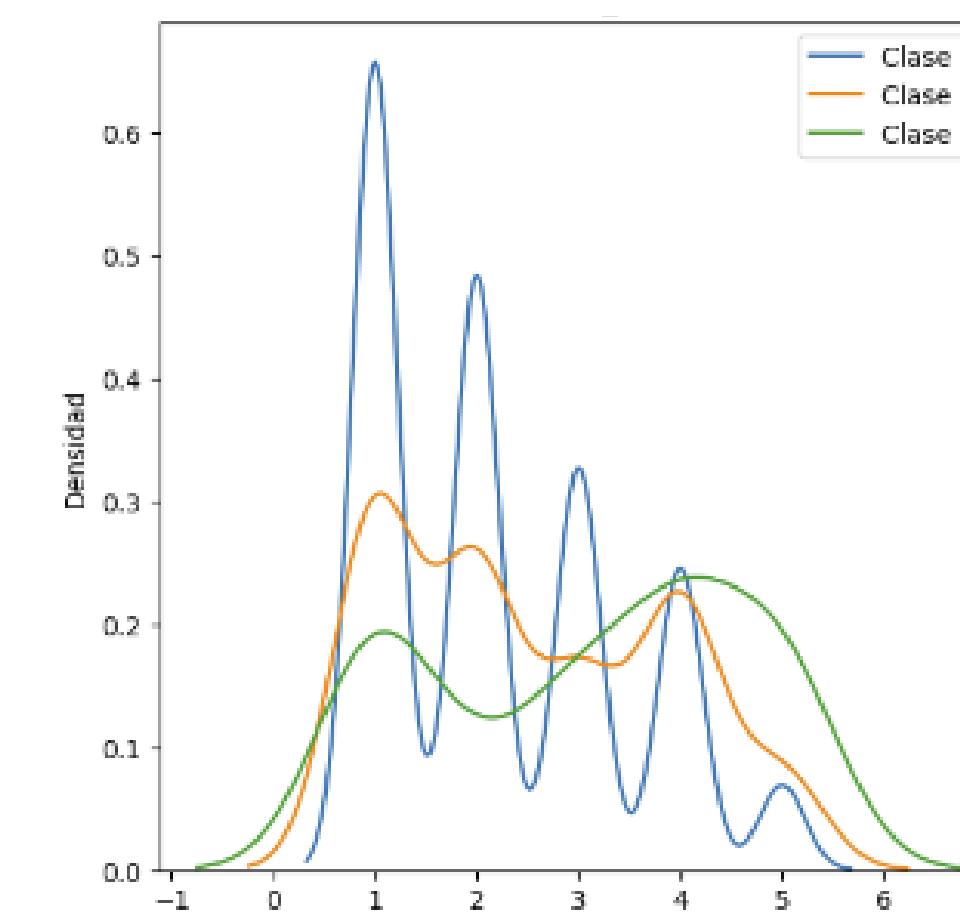
## PLANIFICACION

Distribución de planificación por clase



## PREDIC\_PREMIOS

Distribución de predicción de premios por clase



# MODELOS

---

1. KNN
2. XGBOOST
3. SVC
4. RANDOM FOREST



**Model: KNN**

	precision	recall	f1-score	support
0	0.89	0.99	0.94	825
1	0.36	0.06	0.11	81
2	1.00	0.11	0.19	28
accuracy			0.89	934
macro avg	0.75	0.39	0.41	934
weighted avg	0.85	0.89	0.85	934

**Model: SVC**

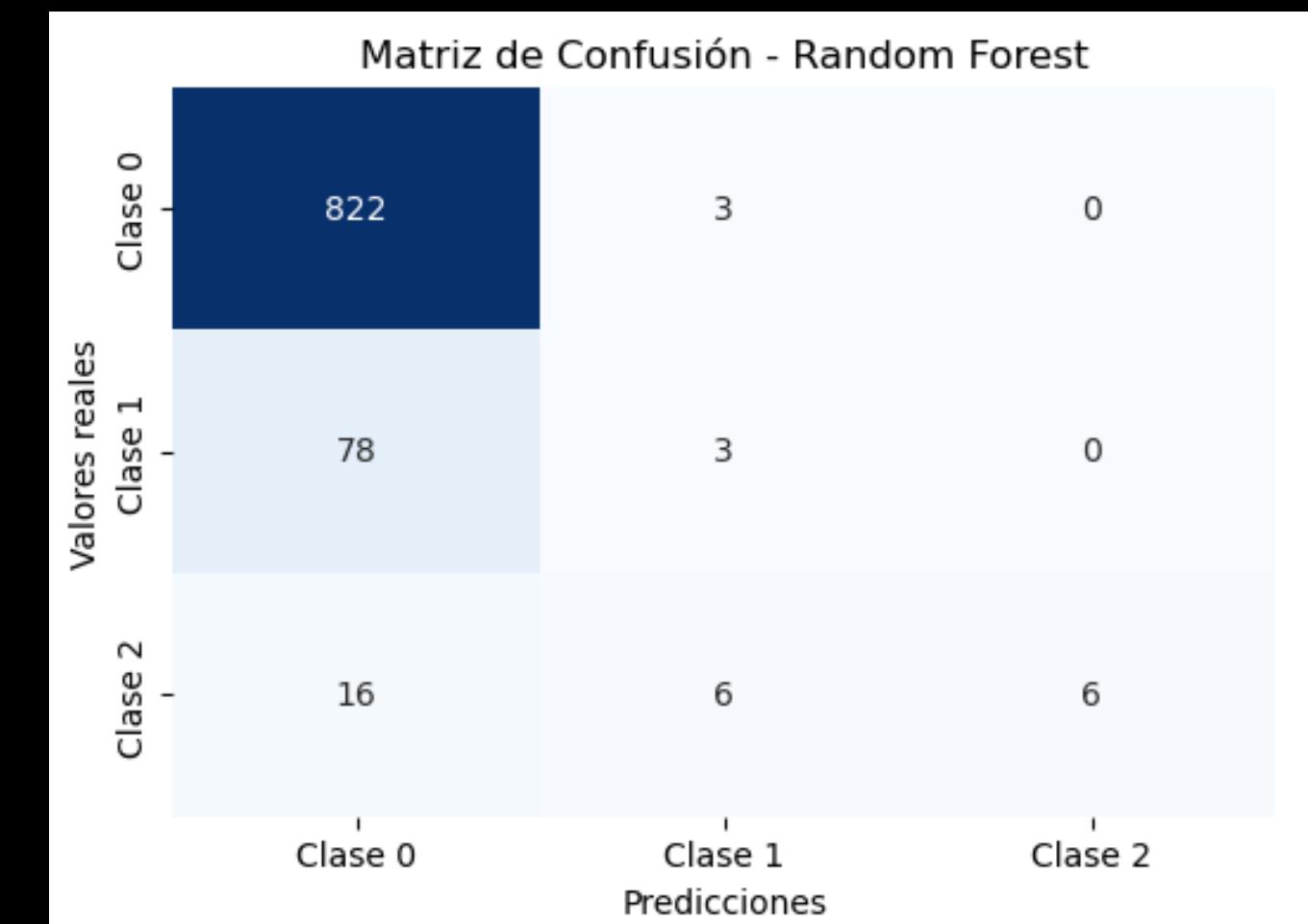
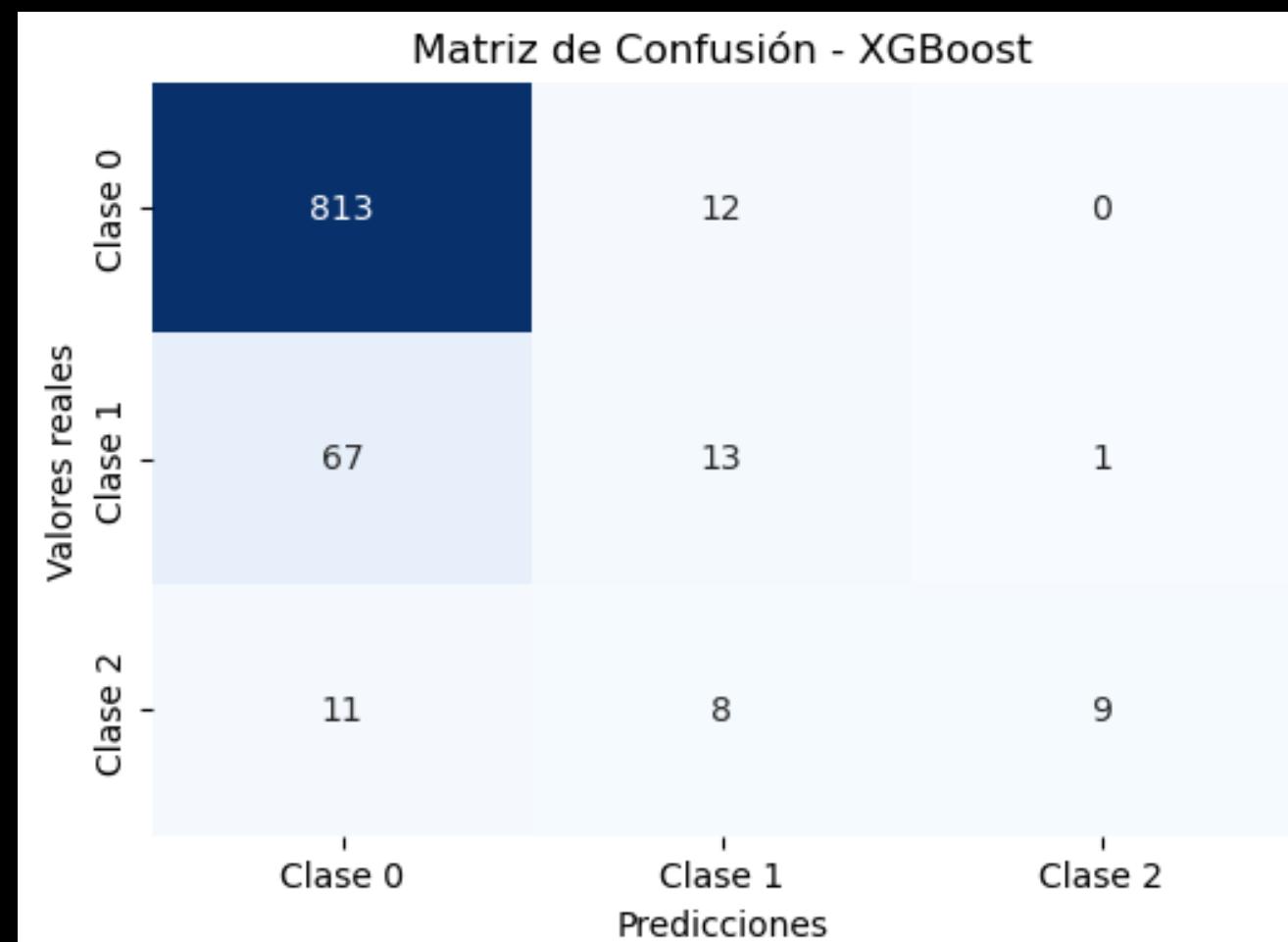
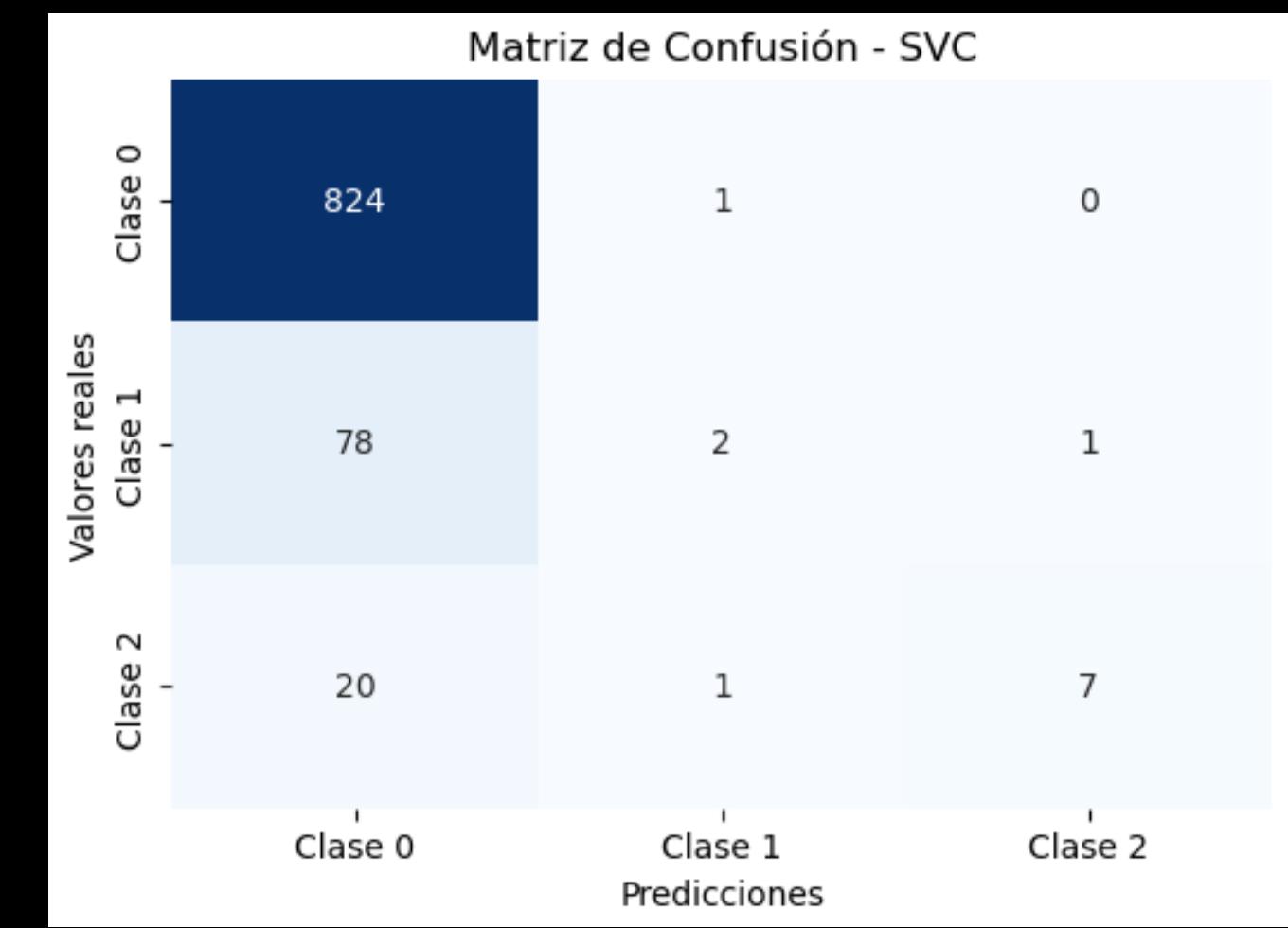
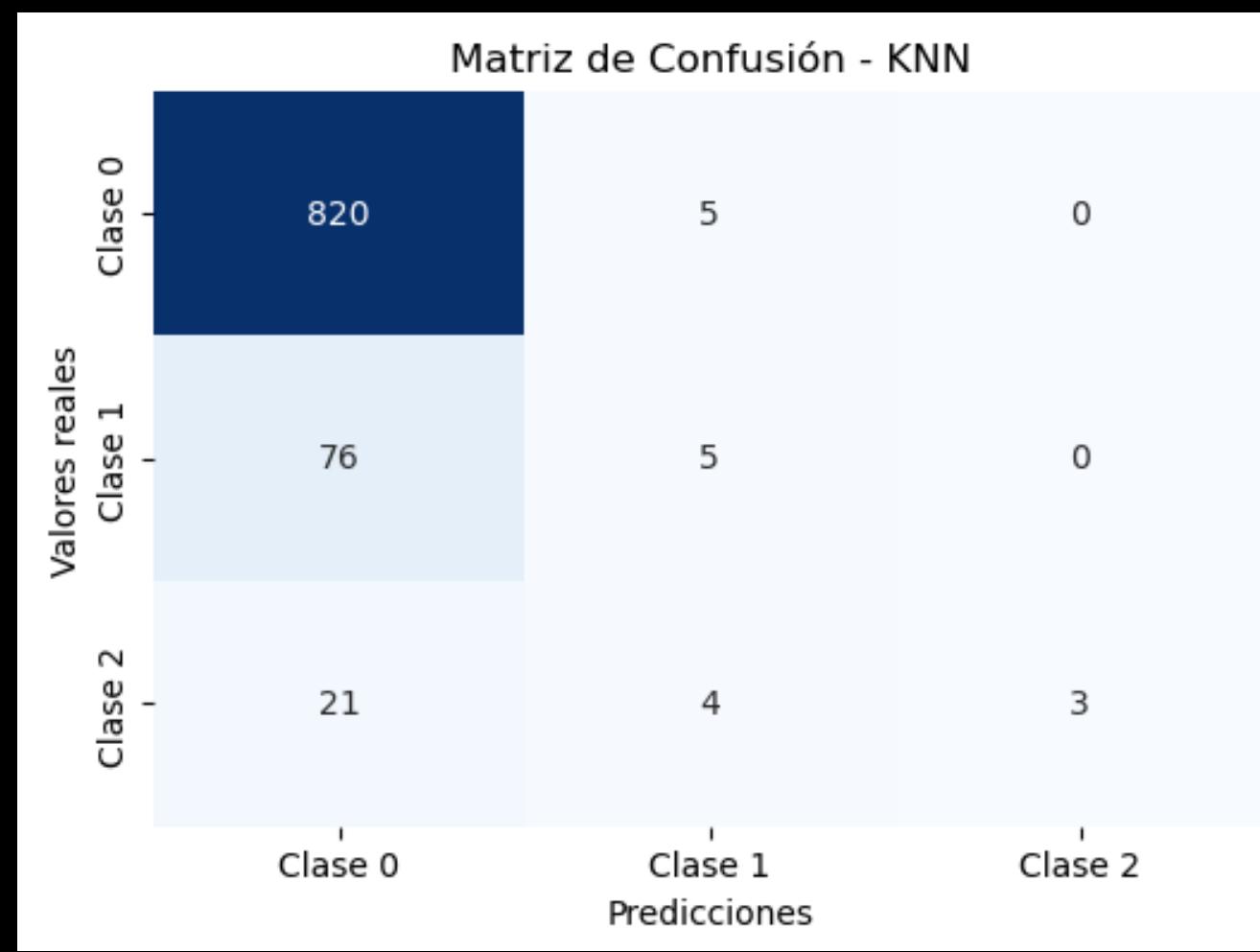
	precision	recall	f1-score	support
0	0.89	1.00	0.94	825
1	0.50	0.02	0.05	81
2	0.88	0.25	0.39	28
accuracy			0.89	934
macro avg	0.76	0.42	0.46	934
weighted avg	0.86	0.89	0.85	934

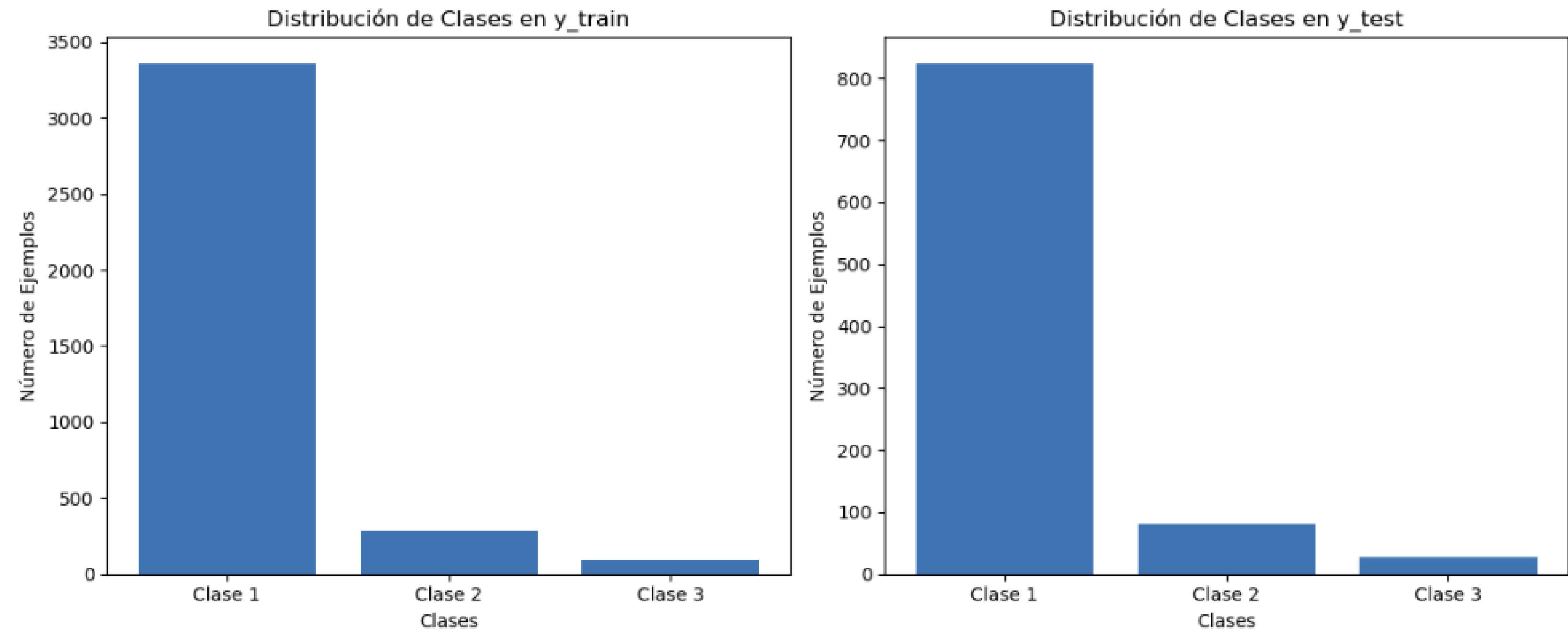
**Model: XGBoost**

	precision	recall	f1-score	support
0	0.91	0.99	0.95	825
1	0.39	0.16	0.23	81
2	0.90	0.32	0.47	28
accuracy			0.89	934
macro avg	0.74	0.49	0.55	934
weighted avg	0.87	0.89	0.87	934

**Model: Random Forest**

	precision	recall	f1-score	support
0	0.90	1.00	0.94	825
1	0.25	0.04	0.06	81
2	1.00	0.21	0.35	28
accuracy			0.89	934
macro avg	0.72	0.42	0.45	934
weighted avg	0.84	0.89	0.85	934





**Clase 0: 4187**

**Clase 1: 363**

**Clase 2: 119**

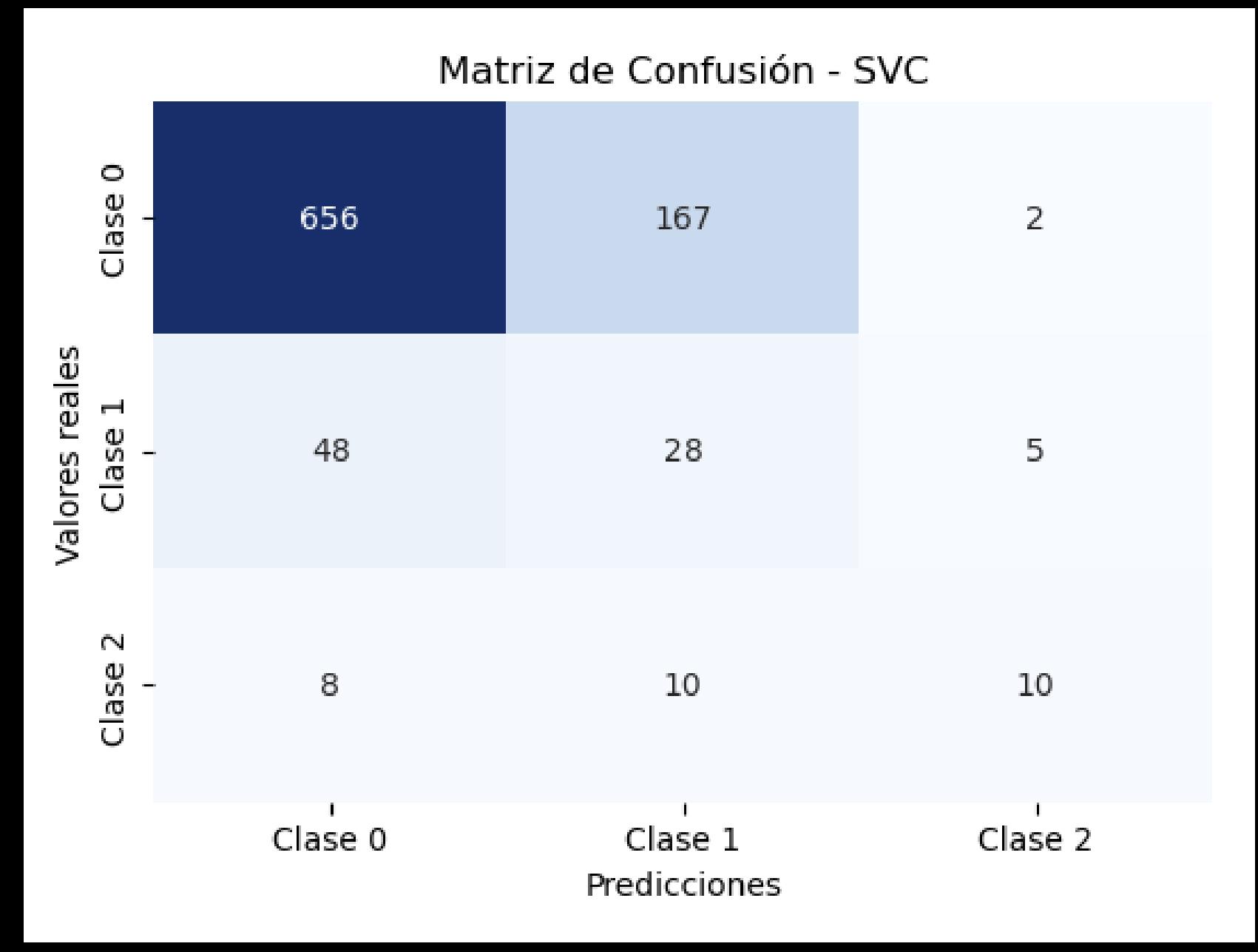
# PONDERACIÓN DE CLASES

⚠️ ATENCIÓN SPOILER: no funciona

Model: SVC

	precision	recall	f1-score	support
0	0.92	0.80	0.85	825
1	0.14	0.35	0.20	81
2	0.59	0.36	0.44	28
accuracy			0.74	934
macro avg	0.55	0.50	0.50	934
weighted avg	0.84	0.74	0.78	934

class\_weights = {0: 1, 1: 100, 2: 300}



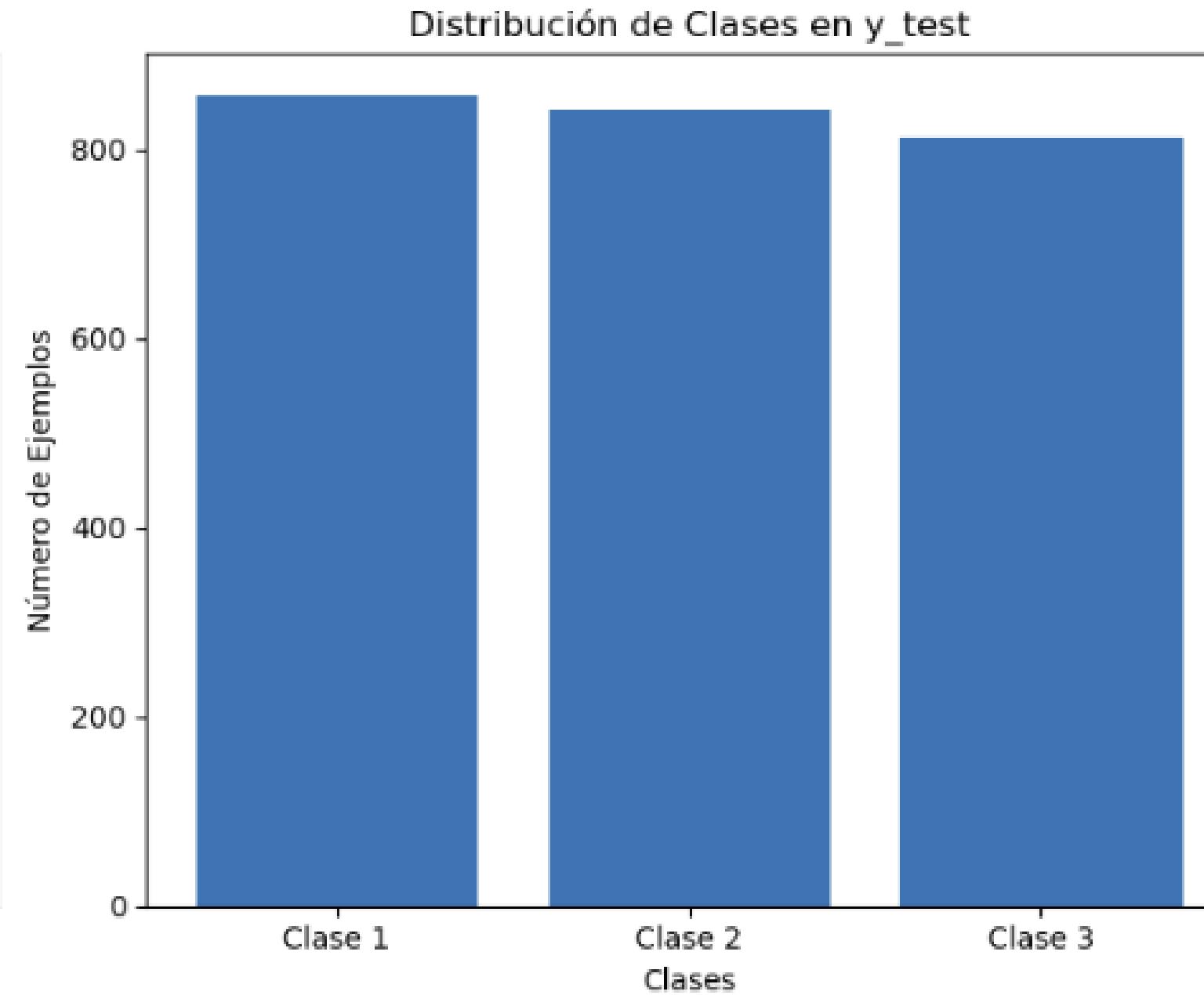
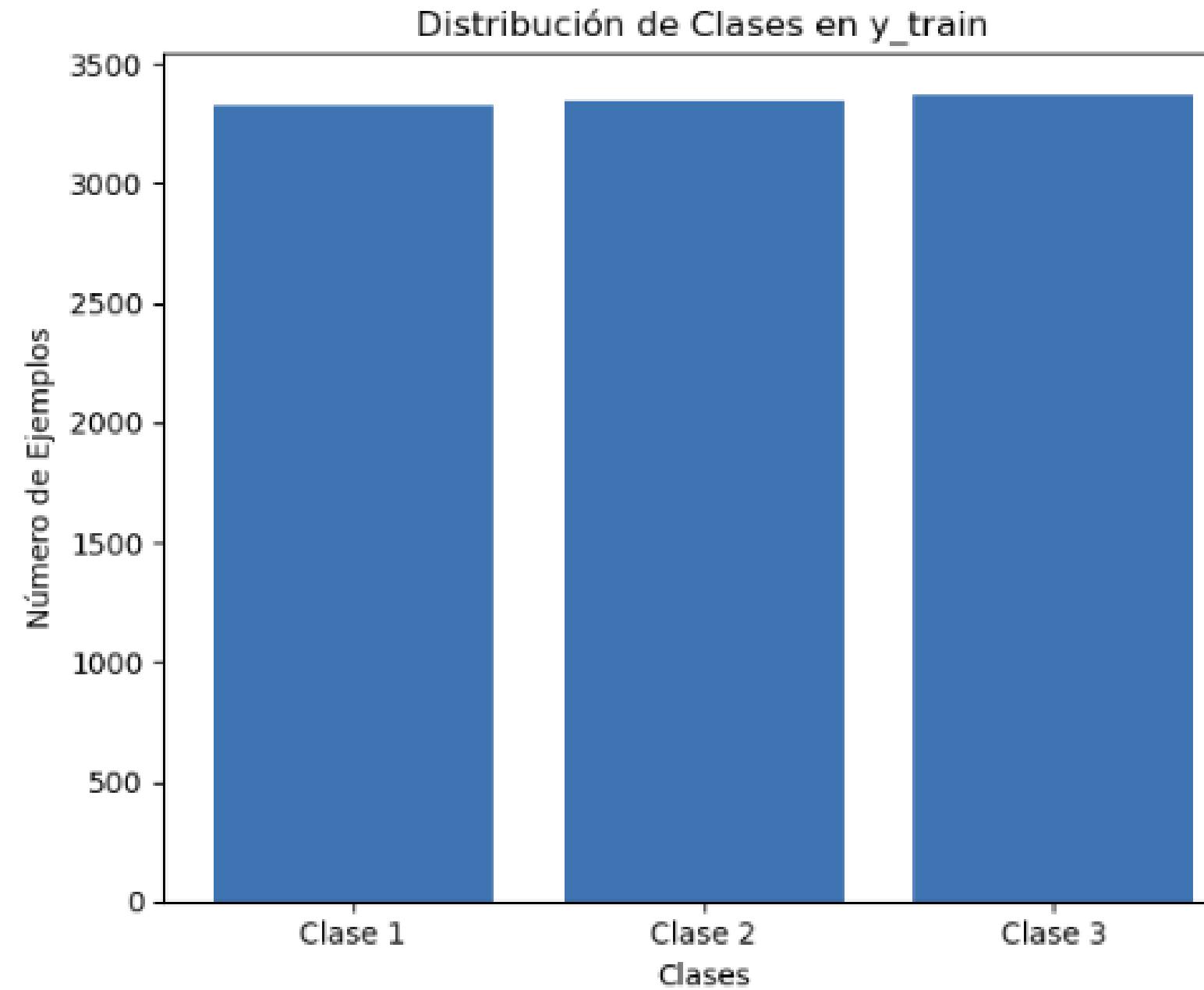
# SMOTE

Sobremuestreo de la clase minoritaria

```
smote = SMOTE(sampling_strategy='auto', random_state=42)
```

```
X_resampled, y_resampled = smote.fit_resample(X, y)
```

```
X_train, X_test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.2)
```



**Model: KNN**

	precision	recall	f1-score	support
0	1.00	0.75	0.86	858
1	0.81	1.00	0.90	842
2	0.97	1.00	0.98	813
accuracy			0.91	2513
macro avg	0.93	0.92	0.91	2513
weighted avg	0.93	0.91	0.91	2513

**Model: SVC**

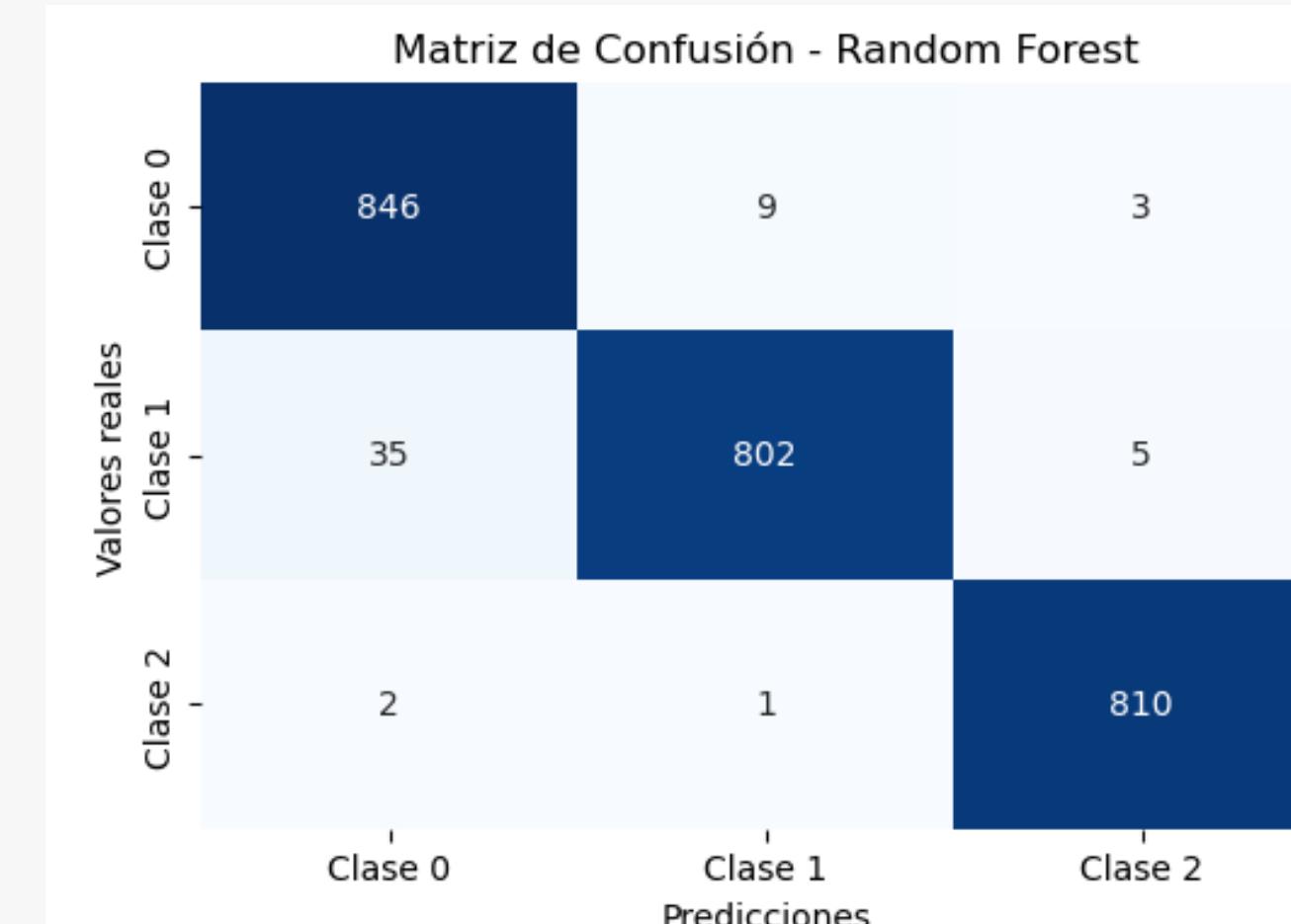
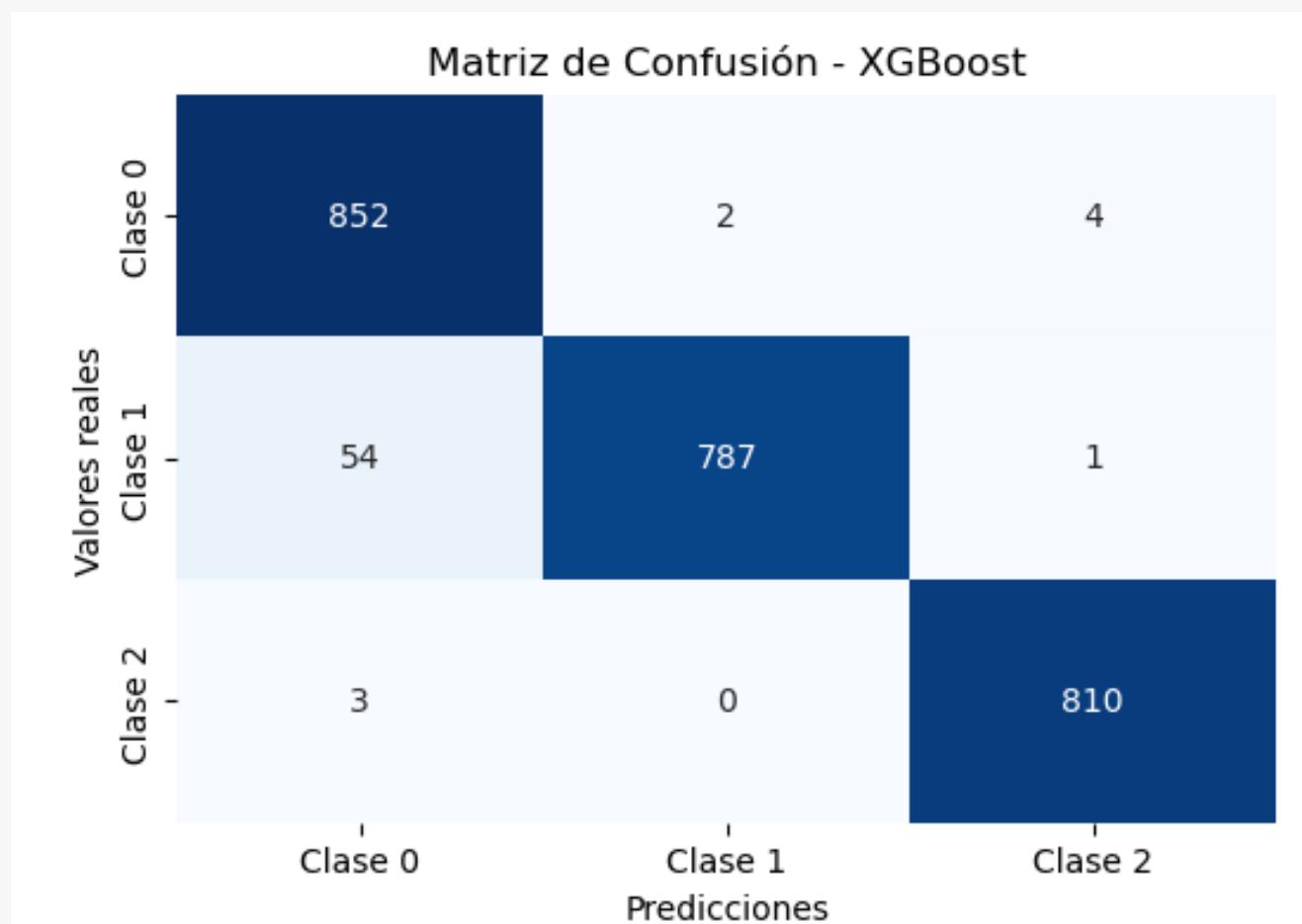
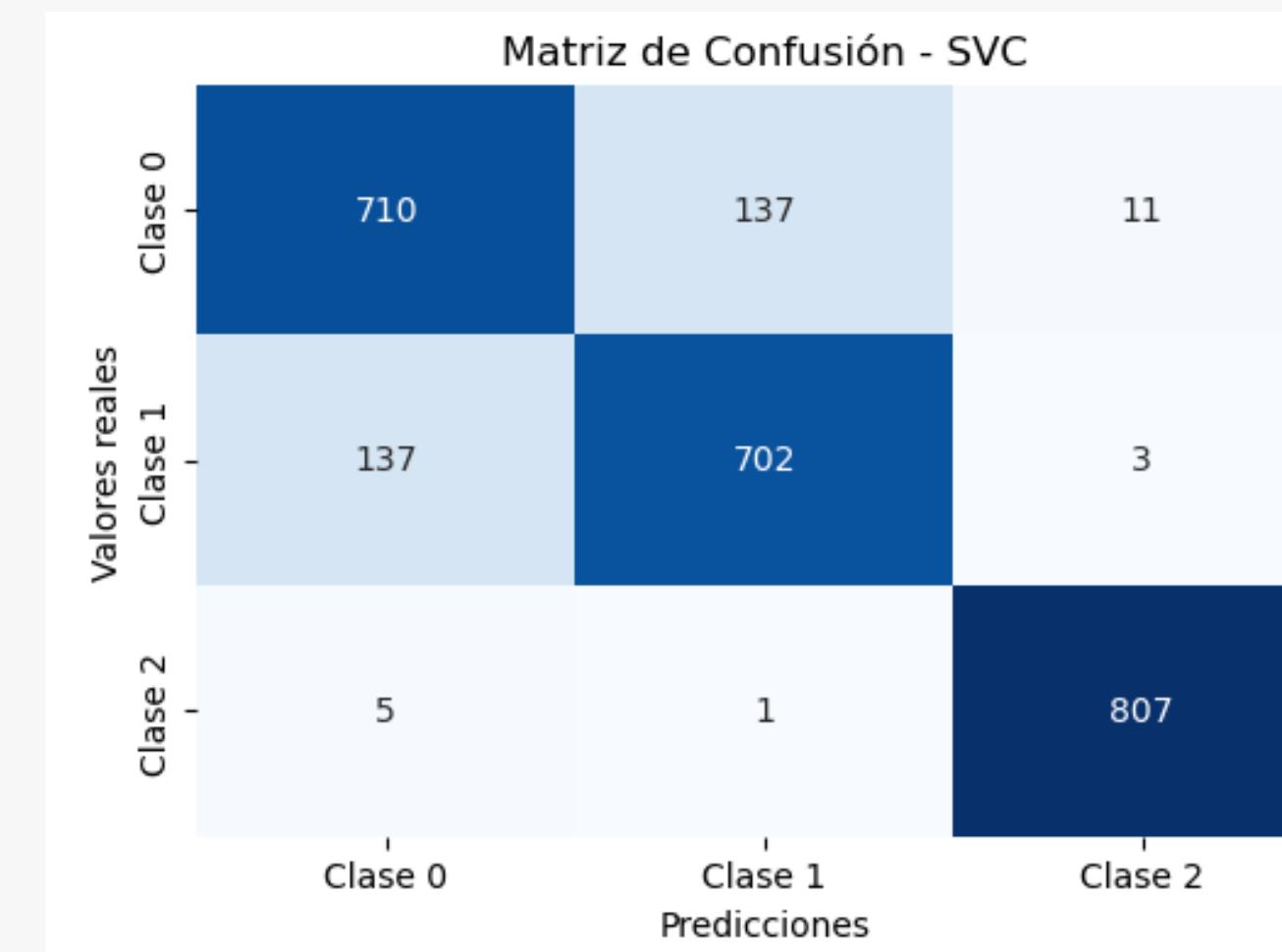
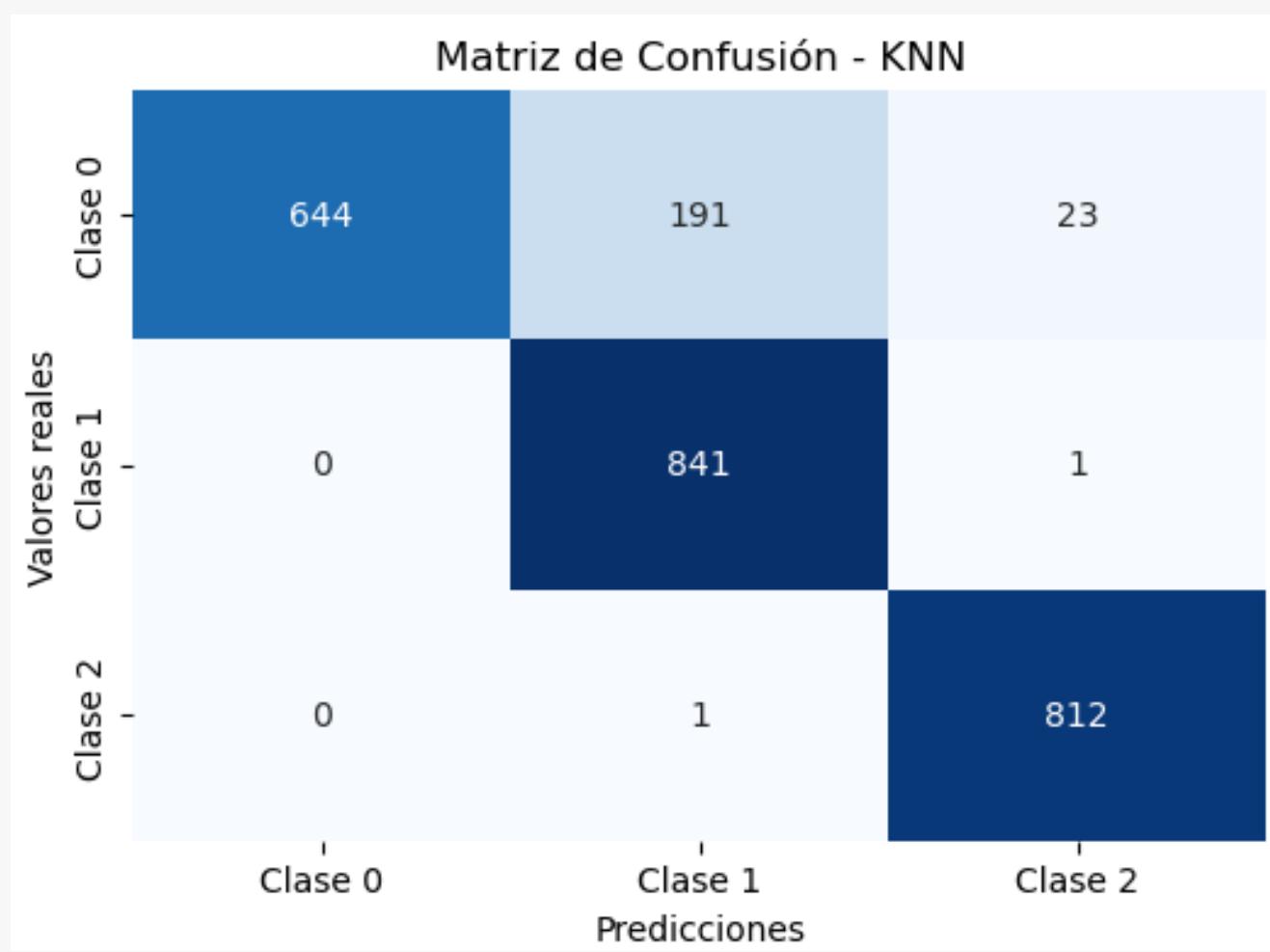
	precision	recall	f1-score	support
0	0	0.83	0.83	0.83
1	0	0.84	0.83	0.83
2	0	0.98	0.99	0.99
accuracy				0.88
macro avg		0.88	0.88	0.88
weighted avg		0.88	0.88	0.88

**Model: XGBoost**

	precision	recall	f1-score	support
0	0.94	0.99	0.96	858
1	1.00	0.93	0.97	842
2	0.99	1.00	1.00	813
accuracy			0.97	2513
macro avg	0.98	0.97	0.97	2513
weighted avg	0.98	0.97	0.97	2513

**Model: Random Forest**

	precision	recall	f1-score	support
0	0	0.96	0.99	0.97
1	0	0.99	0.95	0.97
2	0	0.99	1.00	0.99
accuracy				0.98
macro avg		0.98	0.98	0.98
weighted avg		0.98	0.98	0.98



# PERO NO ES ORO TODO LO QUE RELUCE: CROSS VALIDATION

<b>Modelo: Random Forest</b>	<b>precision</b>	<b>recall</b>	<b>f1-score</b>	<b>support</b>
<b>Clase 0</b>	0.91	1.00	0.95	4187
<b>Clase 1</b>	0.47	0.35	0.39	363
<b>Clase 2</b>	0.70	0.63	0.65	119
<b>macro avg</b>	0.69	0.65	0.65	4669
<b>weighted avg</b>	0.87	0.89	0.85	4669

# CONCLUSIONES

---

**Necesidad de encontrar un buen dataset**

**Finalidad del dataset:  
análisis de datos  
vs  
modelaje**

**Seguimiento del estudio**

**Supervisión por parte de profesionales del ámbito de la salud**



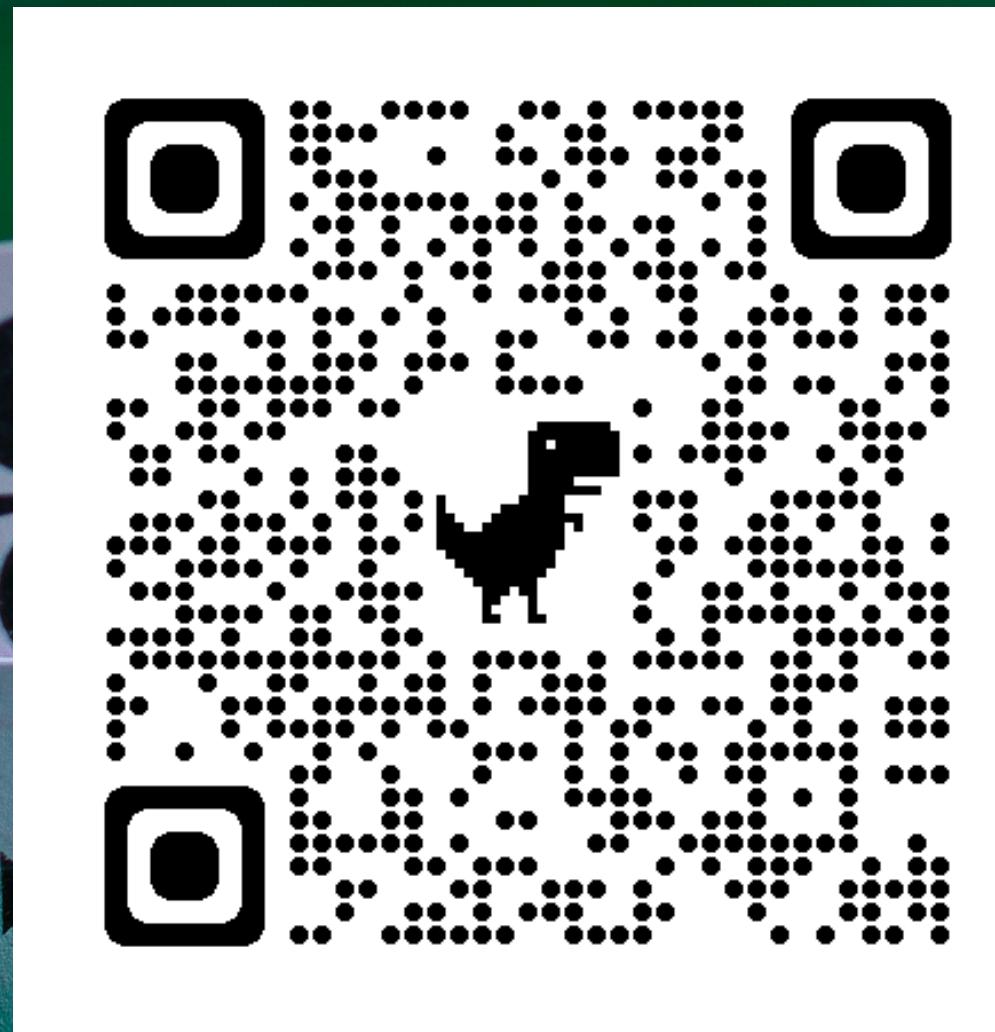


**GRACIAS!**

# CONTINUARÁ...

---

**GITHUB** @dscorbella



**LINKEDIN** @dscorbella



- P14A1 ¿Ha tenido períodos de 2 o más semanas en las que pasase una gran cantidad de tiempo pensando en sus experiencias con el juego o planificando detalladamente futuros episodios de juego o de apuestas?
- P14A2 ¿Ha tenido períodos de 2 o más semanas en las que pasase mucho tiempo pensando en cómo conseguir dinero para jugar?
- P14A3 ¿Ha tenido períodos de 2 o más semanas en los que necesitaba jugar con cantidades de dinero cada vez mayores, o apuestas mayores que antes, para conseguir la misma excitación?
- P14A4 ¿Ha intentado alguna vez dejar, reducir o controlar su juego?
- P14A5 En una o más de estas ocasiones de intento de dejar, reducir o controlar su juego, ¿se sintió intranquilo o irritable?
- P14A6 ¿Alguna vez ha intentado dejar, reducir o controlar su conducta de juego sin poder conseguirlo?
- P14A7 En el caso de que fuese así, ¿ha sucedido 3 o más veces?
- P14A8 ¿Ha jugado alguna vez como una forma de escapar de los problemas personales?
- P14A9 ¿Ha jugado alguna vez para aliviar sentimientos desagradables como culpabilidad, ansiedad, indefensión o depresión?
- P14A10 ¿Ha tenido alguna vez un período en el cual si perdía dinero en el juego volvía otro día para recuperarlo?
- P14A11 ¿Ha mentido alguna vez a su familia, amigos o a otros sobre cuánto juega o cuánto dinero perdía en el juego?
- P14A12 Si es así, ¿esto ha sucedido 3 o más veces?
- P14A13 ¿Ha extendido alguna vez un cheque sin fondos o cogido dinero que no era suyo de familiares u otra persona para gastar en el juego?
- P14A14 ¿Le ha causado alguna vez el juego problemas graves o repetidos en su relación con algún familiar o amigo?
- P14A15 ¿Le ha producido el juego algún problema con los estudios, como por ejemplo perder clases o días de escuela o suspender algún curso?
- P14A16 ¿Le ha causado el juego la pérdida de un trabajo, tener problemas en el trabajo o no poder aprovechar una oportunidad profesional importante?
- P14A17 ¿Ha necesitado alguna vez pedir dinero prestado a un familiar, o a otra persona, para poder salir de una situación económica desesperada causada en gran parte por el juego?