

Royal Flush

Classification Models For A Poker Hands Dataset

Isabella Callejas Mandon - 2202030

Geiner D. Guevara Vargas - 2201840

F1 - IA I

**Universidad Industrial de Santander
2023-2**

CONTENIDO

01. Problema

02. Dataset

03. Métodos de clasificación

04. Reducción de dimensión

05. Resultados

06. Conclusiones

PROBLEMA

CONTEXTO



En póquer, existen combinaciones específicas de cartas, unas más probables que otras y dependiendo del tipo de juego pueden o no beneficiar a los jugadores.

PROBLEMA

OBJETIVO



(9 - *Royal Flush*)

**“Clasificar manos de póquer según
la combinación que presenten las cartas”**

DATASET

ABSTRACCIÓN

En una baraja estándar, cada una de las 52 cartas tiene dos atributos:

1. **Palo o suit (S):** Corazón, Trébol, Diamante, Picas. Representación respectiva [1-4]
2. **Rango (C):** Ace (A), 2-10, Jack o Joker (J), Queen (Q), King (K). Representación respectiva [1-13].



Jack de Corazones

S=1

C=11

En este proyecto no se tomó ningún puntaje, si no la combinación de cartas que se presente en la **mano**.

DATASET

ABSTRACCIÓN

Una **mano de póquer** consiste en 5 cartas, las combinaciones que se van a clasificar según el dataset son las siguientes:

- [0] **Nada en mano:** No entra en las otras clasificaciones
- [1] **Un par:** Dos cartas con el mismo rango, diferente palo
- [2] **Dos pares:** Dos pares de cartas, los pares son diferentes entre sí
- [3] **Trío:** 3 cartas con el mismo rango, diferente palo
- [4] **Escalera:** Las 5 cartas secuenciales sin importar el palo

DATASET

ABSTRACCIÓN

Una **mano de póquer** consiste en 5 cartas, las combinaciones que se van a clasificar según el dataset son las siguientes:

- [5] **Flush/ Color/ Flor:** 5 cartas del mismo palo
- [6] **Full House:** Trio + Par
- [7] **Cuarto de un tipo:** 4 cartas de un mismo rango y diferentes palos.
- [8] **Escalera de color:** Escalera pero todas son del mismo palo
- [9] **Flor Imperial / Royal Flush:** conformado por: A, K, Q, J, 10 y todas son del mismo palo

DATASET

FUENTE

	S1	C1	S2	C2	S3	C3	S4	C4	S5	C5	Class
0	3	12	3	2	3	11	4	5	2	5	1
1	1	9	4	6	1	4	3	2	3	9	1
2	1	4	3	13	2	13	2	1	3	6	1
3	3	10	2	7	1	2	2	11	4	9	0
4	1	3	4	5	3	4	1	12	4	6	0
...
1025004	3	9	2	6	4	11	4	12	2	4	0
1025005	4	1	4	10	3	13	3	4	1	10	1
1025006	2	1	2	10	4	4	4	1	4	13	1
1025007	2	12	4	3	1	10	1	12	4	9	1
1025008	1	7	3	11	3	3	4	8	3	7	1
1025009 rows × 11 columns											

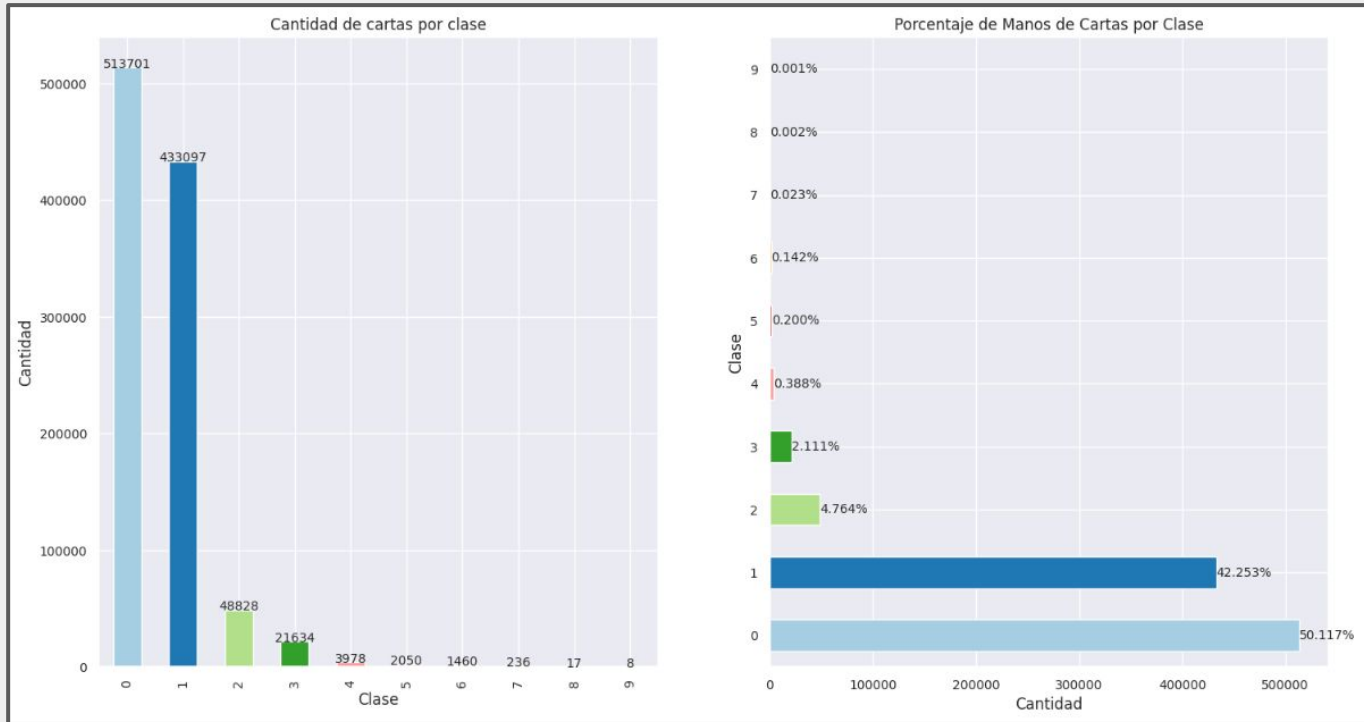
El dataset ya se está hecho y se encuentra disponible en [KEEL](#).

Cuenta con 1'025.009 datos en total

Pero, ¿Cuántos datos realmente tenemos por clase para realizar clasificaciones?

DATASET

PRE-PROCESAMIENTO



Aquí nos enfrentamos a un DESBALANCE de CLASES

DATASET

PRE-PROCESAMIENTO: BALANCE

	S1	C1	S2	C2	S3	C3	S4	C4	S5	C5	Class
3111	2	13	4	10	2	10	1	13	1	10	6
2912	4	11	1	11	3	7	4	7	2	11	6
366	2	9	3	10	1	6	3	11	1	1	0
1423	3	10	4	2	3	9	3	2	4	4	1
3375	4	13	1	13	2	8	2	13	3	13	7
...
2817	2	12	2	5	2	1	2	7	2	9	5
3746	2	3	3	3	1	3	2	9	4	3	7
1601	1	9	4	11	2	5	3	11	2	11	3
1427	4	8	4	6	4	11	3	5	3	11	1
3138	4	13	3	13	1	13	4	12	3	12	6
4800 rows x 11 columns											

Nuestra solución consiste en aumentar las clases más pequeñas con algoritmos que generen 480 combinaciones de cartas a las que pertenece cierta clase (480 es el máximo de permutaciones en Royal Flush, la mano más difícil de lograr) y reduciendo las clases más grandes tomando filas al azar.

Logrando, 480 datos x 10 clases = 4800 datos en total

MÉTODOS DE CLASIFICACIÓN

Gaussian Naive Bayes

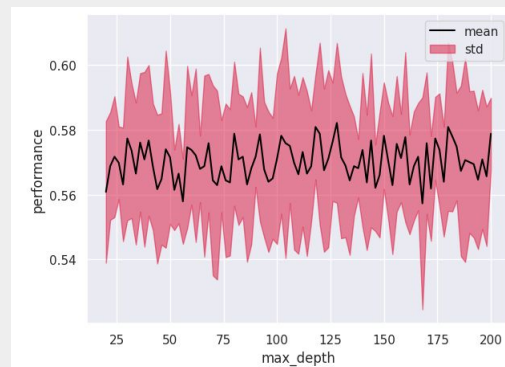
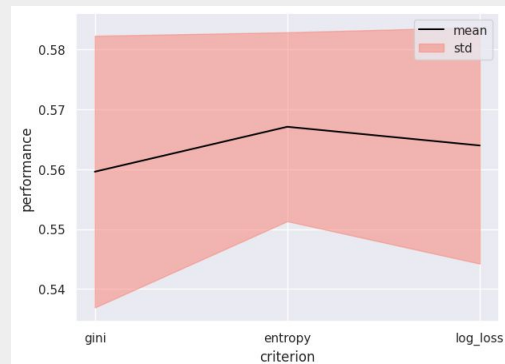
	precision	recall	f1-score	support
0	0.00	0.00	0.00	90
1	0.28	0.80	0.42	206
3	0.00	0.00	0.00	113
4	0.22	0.17	0.19	92
5	0.00	0.00	0.00	110
6	0.00	0.00	0.00	95
7	0.00	0.00	0.00	87
8	0.21	0.18	0.20	87
9	0.34	0.97	0.50	80
accuracy			0.29	960
macro avg	0.12	0.24	0.15	960
weighted avg	0.13	0.29	0.17	960

MÉTODOS DE CLASIFICACIÓN

Decision Tree

El mejor Criterion es entropy y la mejor profundidad es 128

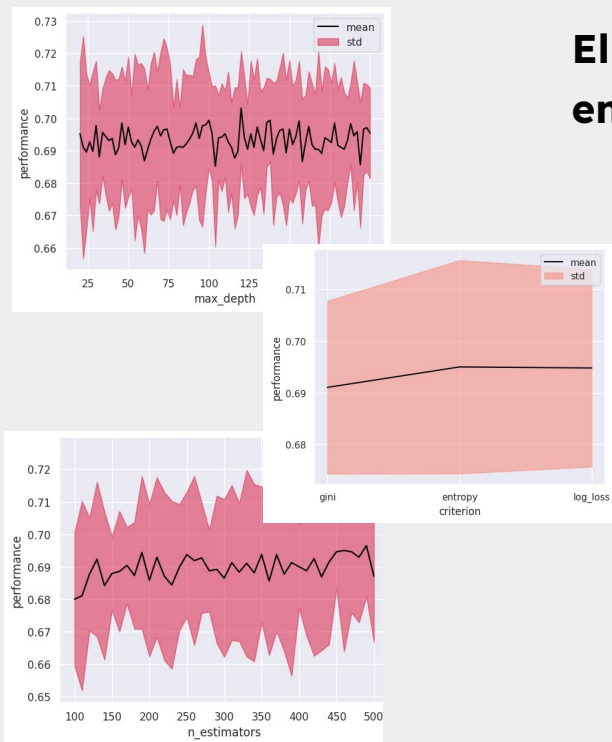
	precision	recall	f1-score	support
0	0.32	0.29	0.30	90
1	0.68	0.85	0.76	206
3	0.27	0.19	0.23	113
4	0.49	0.47	0.48	92
5	0.48	0.40	0.44	110
6	0.36	0.33	0.34	95
7	0.43	0.45	0.44	87
8	0.55	0.61	0.58	87
9	0.94	1.00	0.97	80
accuracy			0.54	960
macro avg	0.50	0.51	0.50	960
weighted avg	0.51	0.54	0.52	960



MÉTODOS DE CLASIFICACIÓN

Random Forest

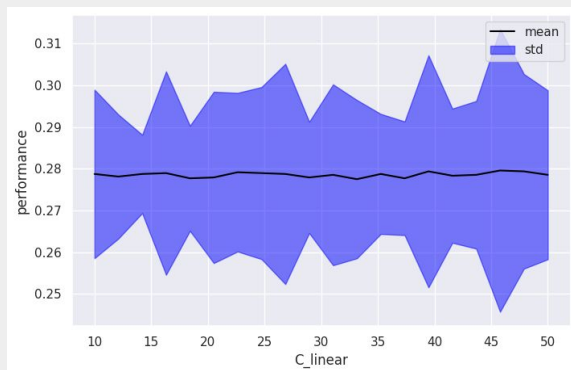
El mejor N. estimadores es 490, Criterion es entropy y la mejor profundidad es 120



	precision	recall	f1-score	support
0	0.51	0.26	0.34	90
1	0.61	0.88	0.72	206
3	0.33	0.11	0.16	113
4	0.60	0.87	0.71	92
5	0.89	0.86	0.88	110
6	0.40	0.22	0.29	95
7	0.52	0.56	0.54	87
8	0.79	0.98	0.88	87
9	0.93	1.00	0.96	80
accuracy			0.65	960
macro avg	0.62	0.64	0.61	960
weighted avg	0.61	0.65	0.61	960

MÉTODOS DE CLASIFICACIÓN

SVM: kernel Lineal



El mejor parámetro C es 45.7

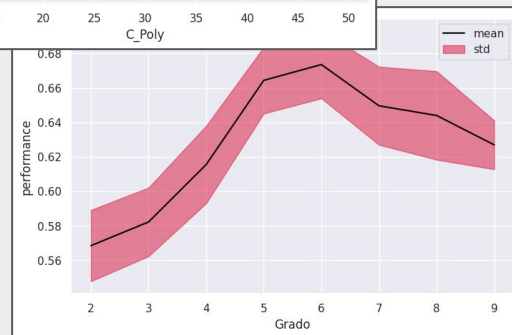
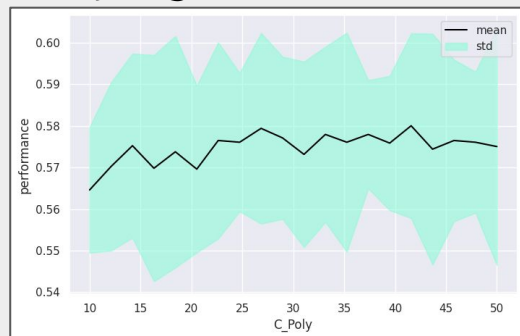
	precision	recall	f1-score	support
0	0.00	0.00	0.00	90
1	0.25	0.89	0.39	206
3	0.00	0.00	0.00	113
4	0.00	0.00	0.00	92
5	0.00	0.00	0.00	110
6	0.00	0.00	0.00	95
7	0.00	0.00	0.00	87
8	0.00	0.00	0.00	87
9	0.37	1.00	0.54	80
accuracy			0.27	960
macro avg	0.07	0.21	0.10	960
weighted avg	0.08	0.27	0.13	960

MÉTODOS DE CLASIFICACIÓN

SVM: kernel Polinomial

El mejor parámetro C es 41.5 y
el mejor grado es 6

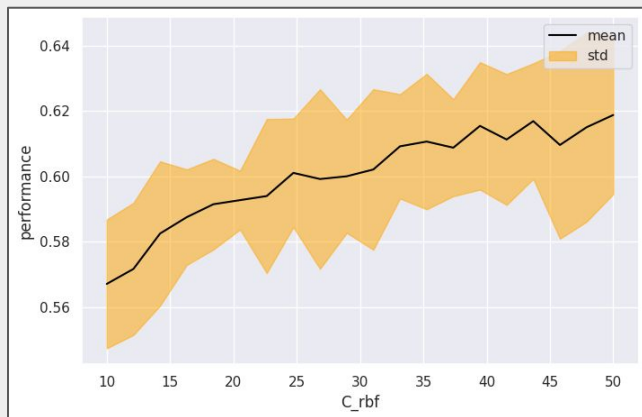
	precision	recall	f1-score	support
0	0.37	0.42	0.39	90
1	0.64	0.82	0.71	206
3	0.27	0.17	0.21	113
4	0.62	0.68	0.65	92
5	0.88	0.86	0.87	110
6	0.47	0.21	0.29	95
7	0.46	0.44	0.45	87
8	0.90	1.00	0.95	87
9	0.88	1.00	0.94	80
accuracy			0.63	960
macro avg	0.61	0.62	0.61	960
weighted avg	0.61	0.63	0.61	960



MÉTODOS DE CLASIFICACIÓN

SVM: kernel función de base radial

El mejor parámetro C es 50



	precision	recall	f1-score	support
0	0.35	0.26	0.29	90
1	0.50	0.63	0.55	206
3	0.31	0.16	0.21	113
4	0.51	0.89	0.65	92
5	0.84	0.85	0.84	110
6	0.32	0.19	0.24	95
7	0.57	0.39	0.46	87
8	0.87	1.00	0.93	87
9	0.91	1.00	0.95	80
accuracy			0.59	960
macro avg	0.57	0.60	0.57	960
weighted avg	0.56	0.59	0.56	960

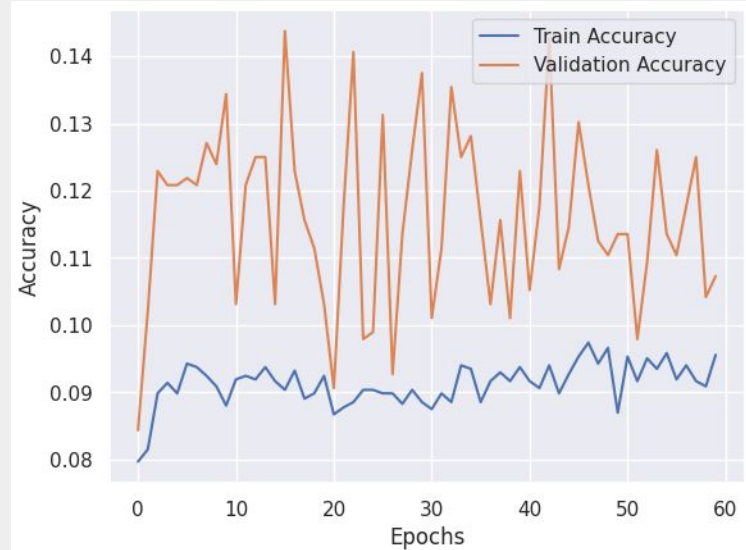
DEEP LEARNING

MODELO DE REDES DENSAS

Model: "sequential_6"

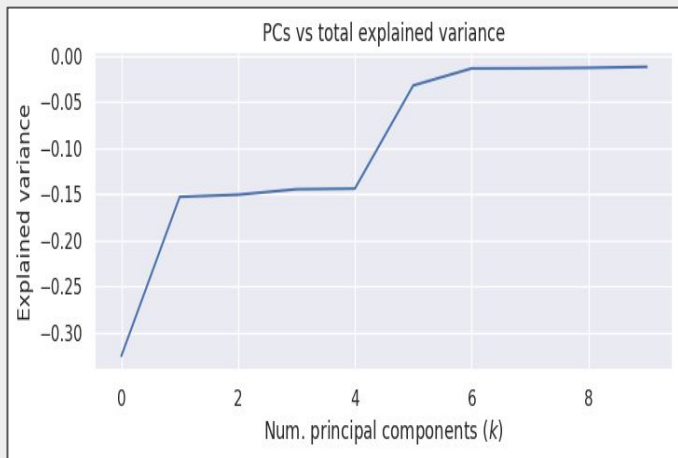
Layer (type)	Output Shape	Param #
flatten_6 (Flatten)	(None, 10)	0
dense_35 (Dense)	(None, 2048)	22528
dense_36 (Dense)	(None, 1024)	2098176
dense_37 (Dense)	(None, 512)	524800
dense_38 (Dense)	(None, 10)	5130

Total params: 2650634 (10.11 MB)
Trainable params: 2650634 (10.11 MB)
Non-trainable params: 0 (0.00 Byte)



REDUCCIÓN DE DIMENSIÓN

PCA



```
total_explained_variance (k=0): 0
total_explained_variance (k=1): 0.3253614329048179
total_explained_variance (k=2): 0.47808871890935095
total_explained_variance (k=3): 0.6285022849303967
total_explained_variance (k=4): 0.7729942536899328
total_explained_variance (k=5): 0.916835367211706
total_explained_variance (k=6): 0.948788144382972
total_explained_variance (k=7): 0.9622170512331562
total_explained_variance (k=8): 0.975474226449265
total_explained_variance (k=9): 0.9882481929691632
total_explained_variance (k=10): 0.9999999999999999
```

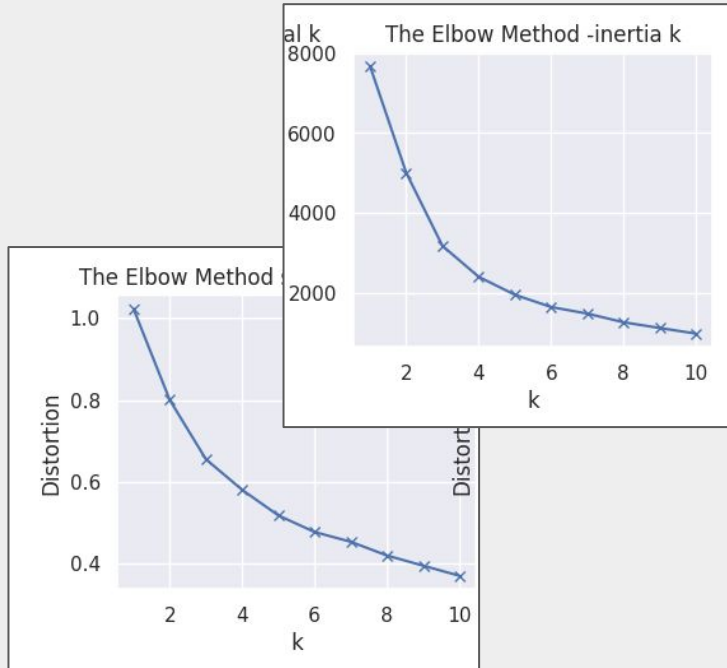
	PCA1	PCA2	PCA3	PCA4	PCA5
0	1.883400	-0.434419	0.157676	-0.490181	-0.677000
1	-0.179413	-0.060158	0.454653	0.241802	-0.003053
2	1.346136	-1.184472	0.543609	0.137508	-0.667426
3	-1.076262	1.067338	-0.370619	-0.219000	-2.340114
4	0.009260	-0.317978	-0.698832	-0.150451	-0.588244
...
4795	0.445735	-1.464477	0.108509	-0.558468	0.617036
4796	-1.261937	-0.178791	0.250479	-0.716061	-0.471130
4797	-0.725754	-1.629611	0.641504	0.242247	0.418481
4798	0.544998	-0.783374	-0.366455	0.387304	-1.018732
4799	0.026685	1.025990	-1.694877	1.785058	1.093826

4800 rows × 5 columns

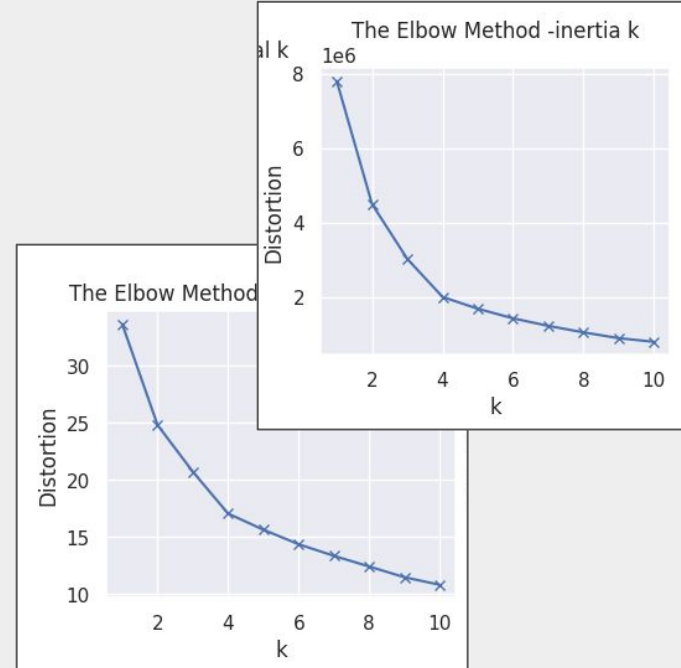
REDUCCIÓN DE DIMENSIÓN

KMEANS

PCA

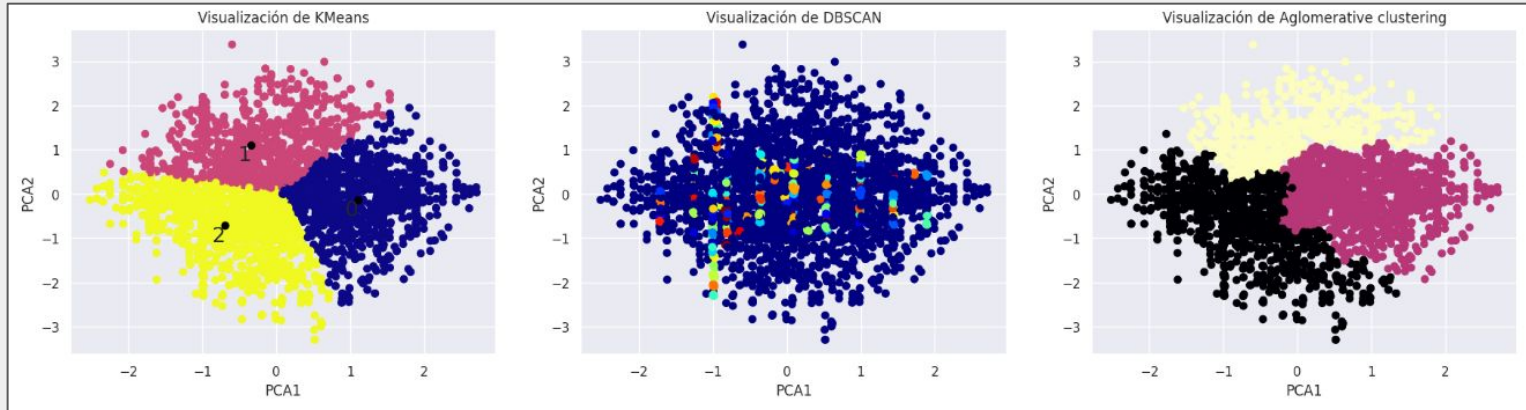


TSN-E



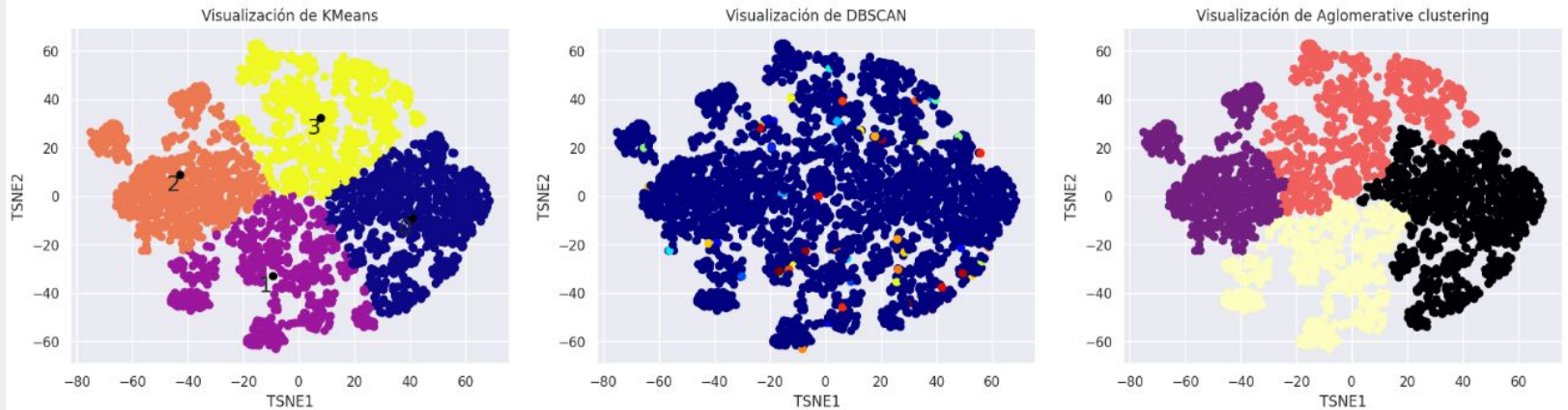
REDUCCIÓN DE DIMENSIÓN

PCA



REDUCCIÓN DE DIMENSIÓN

TSN-E



RESULTADOS

	ACCURACY			
Método	No-Tunning	Tunning	PCA - Tunning	TSN-E - Tunning
GNB	0.30396	0.30125	0.36792	0.32542
Decision Tree	0.57313	0.57188	0.57000	0.57354
Random Forest	0.68354	0.68896	0.61938	0.61229
SVM - Lineal	0.27896	0.27938	0.27500	—
SVM - Polinomial	0.47708	0.67208	0.49896	—
SVM - RBF	0.45333	0.60979	0.50000	—
DNN	0.10729	—	—	—

* El dataset en uso es el que está balanceado.