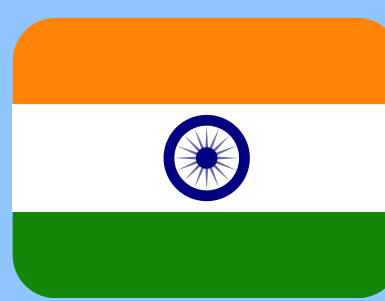
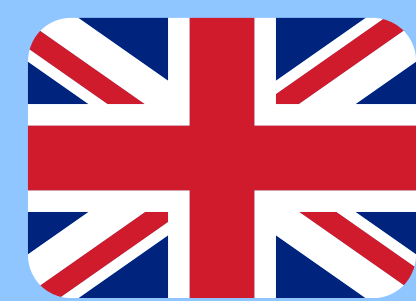
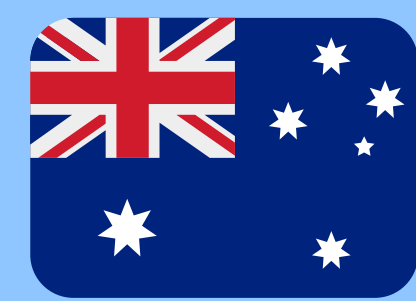




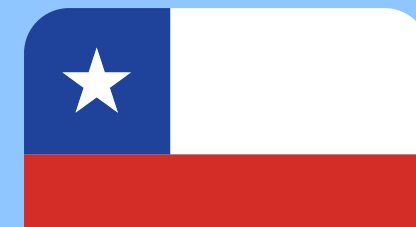
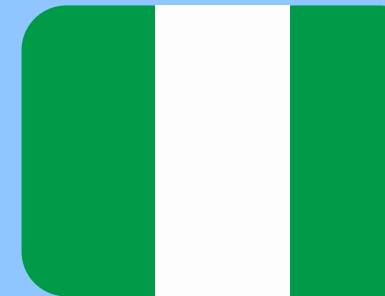
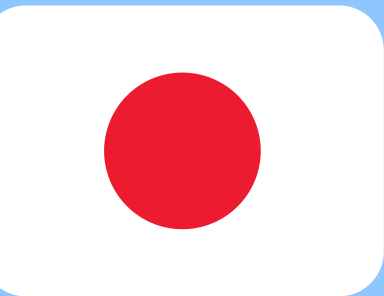
Universidad Nacional  
de San Martín



INTRODUCCIÓN AL APRENDIZAJE AUTOMÁTICO - 2DO CUATRIMESTRE 2023

# ¿QUÉ PODEMOS SABER DE UN PAÍS MIRANDO SU BANDERA?

MARÍA BELEN QUIÑONES, JULIÁN ROLANDO



# Recorrido



Pregunta



Datos



Exploración



Procesado



Métricas y  
modelos



Evaluación



Conclusión

# Datset

## World Flags

<https://www.kaggle.com/datasets/edoardoba/world-flags>

Fuente: Kaggle

**Filas: 193**

**Columnas: 31**

Cada fila representa **un país** y contiene información geográfica y de las **características de la bandera**

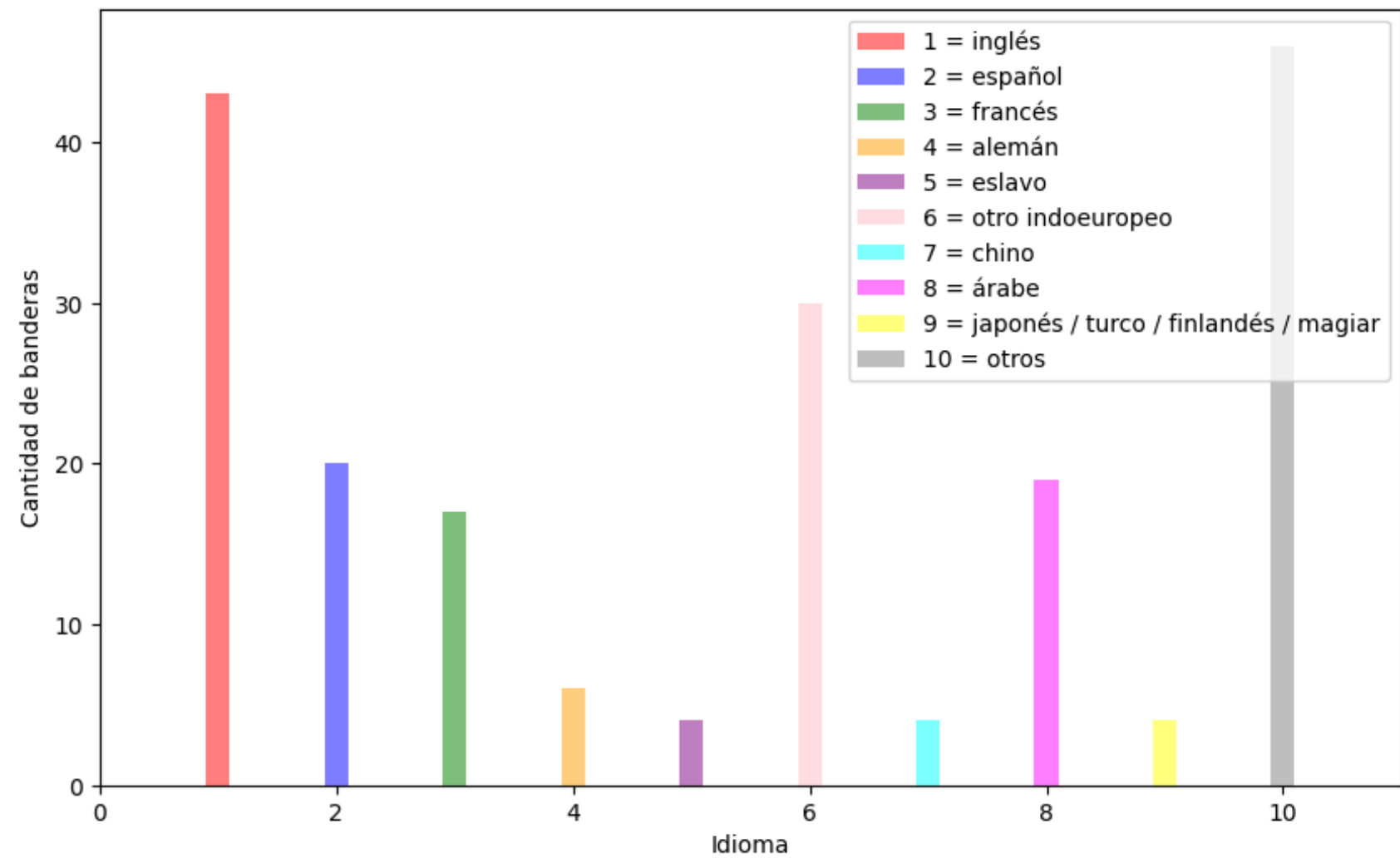
VARIABLE	DESCRIPCIÓN	TIPO
Colores	Cantidad de colores que tiene la bandera	Numerico
Rojo	1 si tiene el rojo, 0 si no lo tiene	Numérico
Verde	1 si tiene el verde, 0 si no lo tiene	Numérico
Azul	1 si tiene el azul, 0 si no lo tiene	Numérico
...	Sucede lo mismo con los colores: dorado blanco, negro y naranja	...
Color principal	Color principal de la bandera	String
opleft	Color en la esquina superior izquierda	String
botright	Color en la esquina inferior izquierda	String

VARIABLE	DESCRIPCIÓN	TIPO
Nombre	Nombre del país	String
Continente	Continente	Numérico
Zona	Cuadrante geográfico	Numérico
Area	Superficie en miles de kilómetros cuadrados	Numérico
Población	Población en millones redondos	Numérico
Idioma	Idioma del país	Numérico
Religión	8 religiones distintas etiquetadas con numeros	Numérico

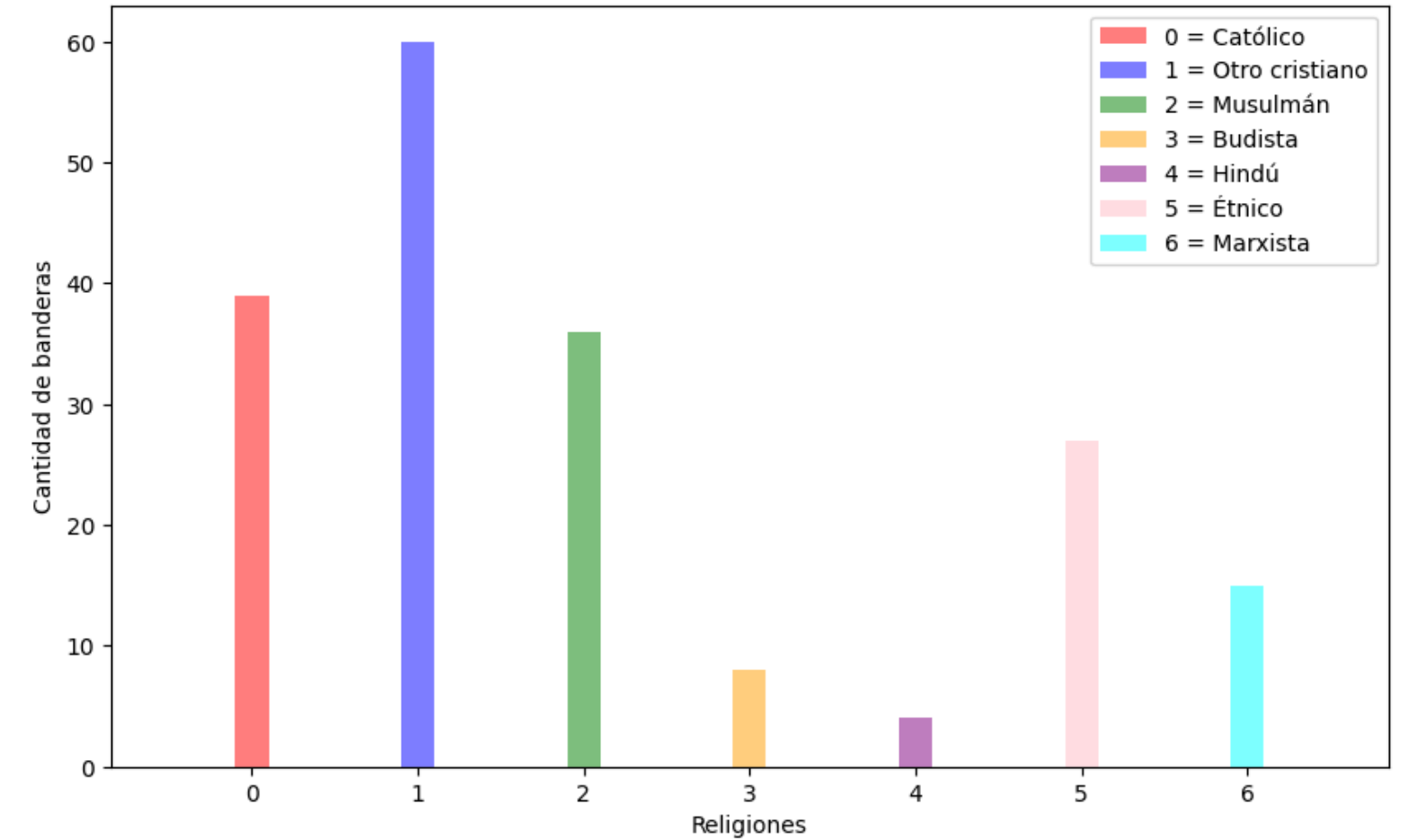
VARIABLE	DESCRIPCIÓN	TIPO
Barras	Cantidad de barras verticales	Numérico
Rayas	Cantidad de rayas horizontales	Numérico
Circulos	Cantidad de circulos	Numérico
Cruces	Cantidad de cruces	Numérico
Cruces diagonales	Cantidad de cruces diagonales	Numérico
Cuartos	Cantidad de cuadrantes	Numérico
Soles	Cantidad de simbolos de soles o estrellas	Numérico
Lunas	1 si tiene el símbolo de una luna creciente, 0 si no	Numérico
Triángulo	1 si tiene algún triangulo, 0 si no	Numérico
Icono	1 si tiene una imagen inanimada, 0 si no	Numérico
Animado	1 si tiene una imagen animada, 0 si no	Numérico
Texto	1 si tiene texto 0 si no	Numérico

# Exploración

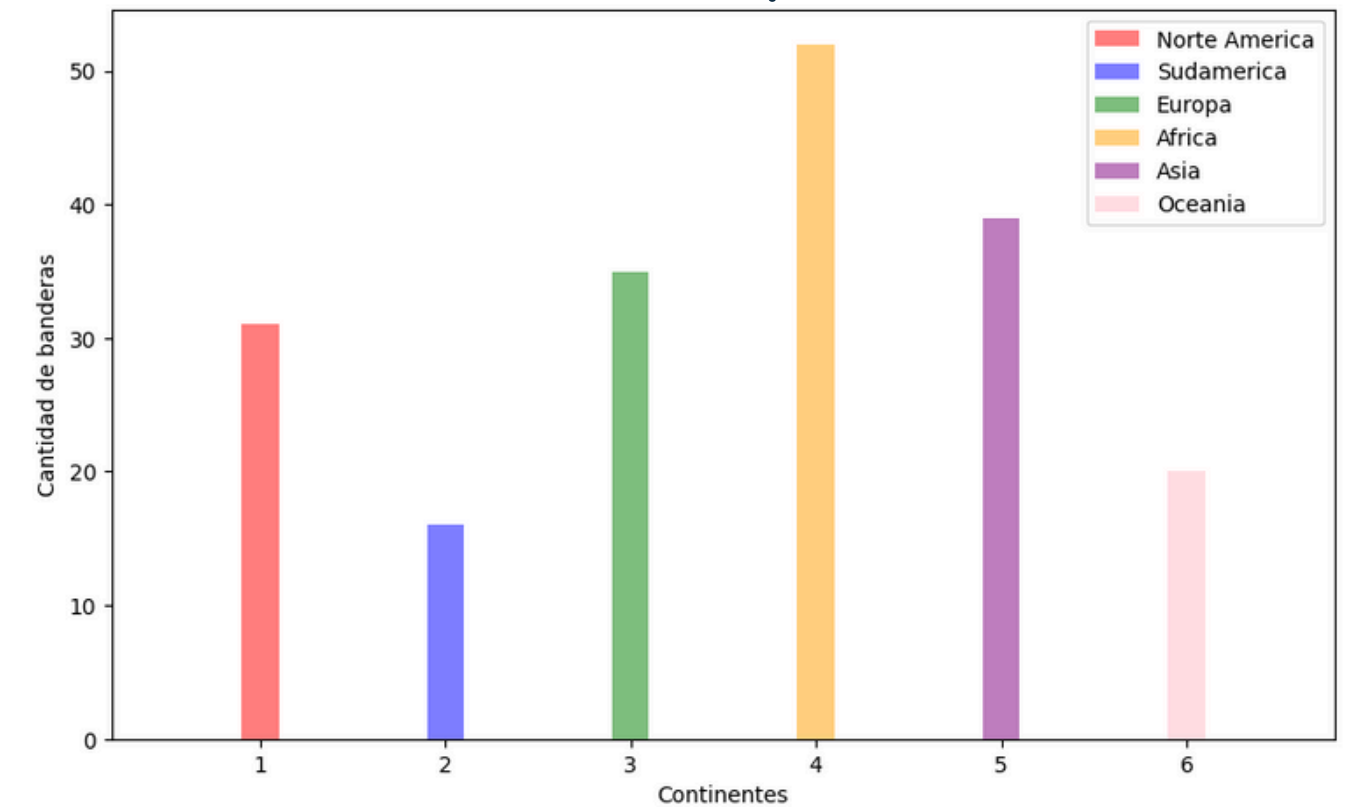
## Cantidad de banderas por idioma



## Cantidad de banderas por religión



## Cantidad de banderas por continente





## Religión

- 0 = Católico
- 1 = Otro cristiano
- 2 = Musulmán
- 3 = Budista
- 4 = Hindú
- 5 = Étnico
- 6 = Marxista
- 7 = Otros



## Continente

- 1 = Norteamérica
- 2 = Sudamérica
- 3 = Europa
- 4 = África
- 5 = Asia
- 6 = Oseanía



## Idioma

- 1 = inglés
- 2 = español
- 3 = francés
- 4 = alemán
- 5 = eslavo
- 6 = otro indoeuropeo
- 7 = chino
- 8 = árabe
- 9 = japonés / turco / finlandés / magiar
- 10 = otros

# Procesado y limpieza

🚩 Cambio del idioma de las columnas

🚩 Descartamos columnas que no nos servían

Topleft, Botright y imagen

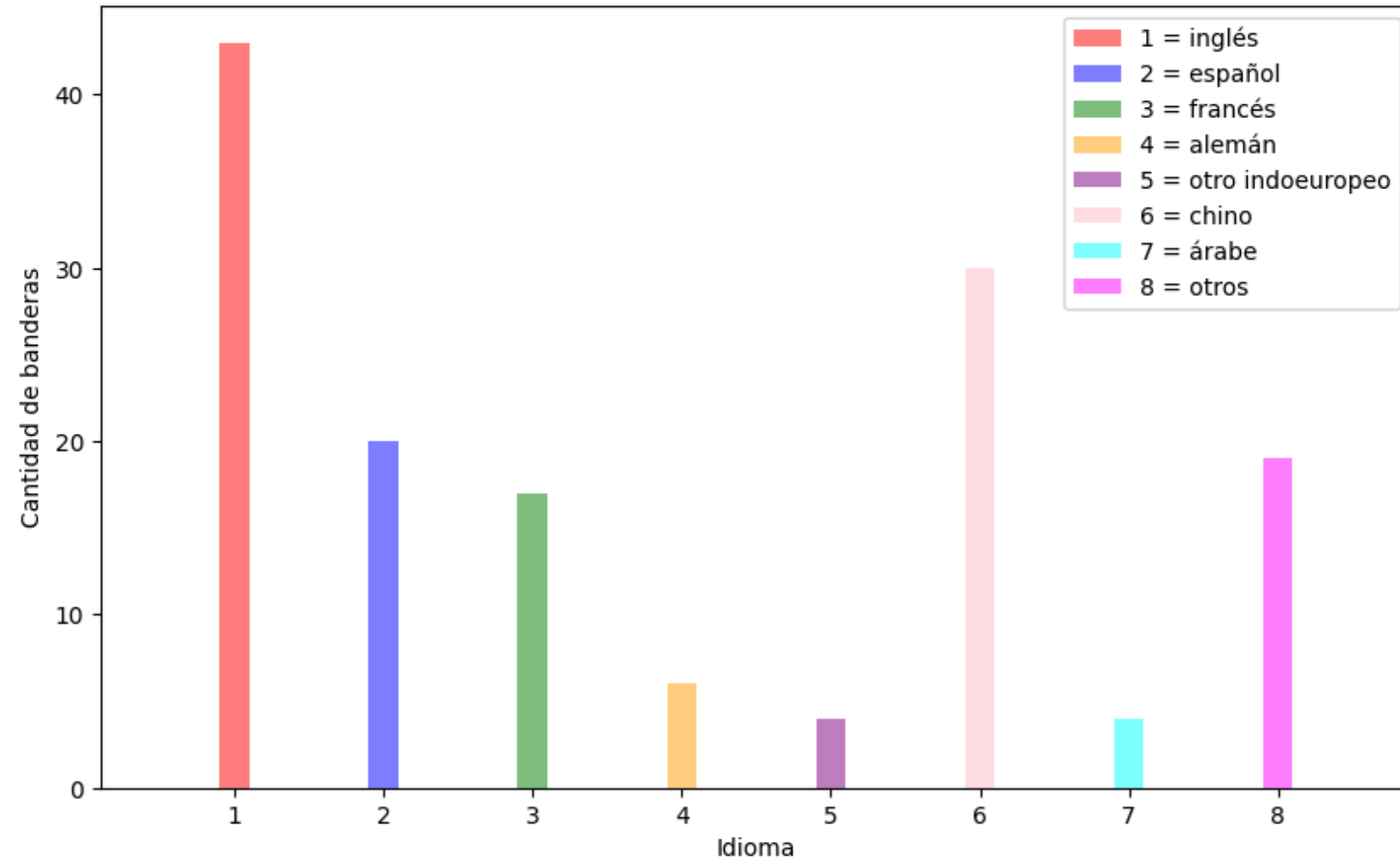
🚩 One Hot Encoding  
Sobre la columna de COLOR PRINCIPAL

🚩 Corrección y simplificación  
Sobre las etiquetas de RELIGIÓN de 19 países y sobre las etiquetas de IDIOMA de 16 países.

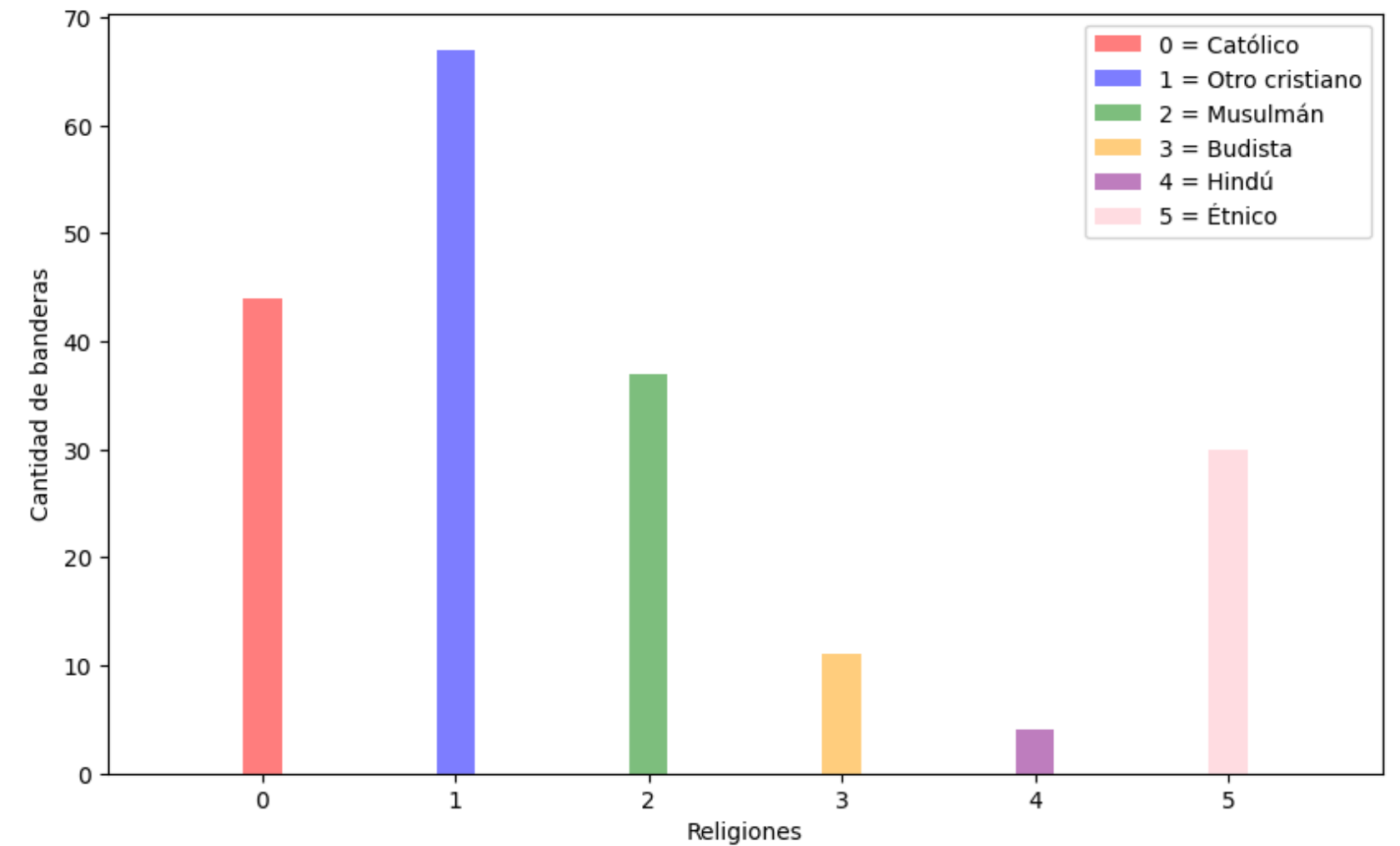
Elegimos corregirlo manualmente

# Nueva distribución

## Cantidad de banderas por idioma



## Cantidad de banderas por religión



# Targets



Religión



Continente



Idioma

# Selección de Features

Colores

ícono

Barras

Animado

Cuartos

Lunas

Rayas

Nos quedamos  
con las  
características  
de las banderas

Texto

Cruces

Círculos

Cruces diagonales

Triángulo

Soles



# Métricas y modelos

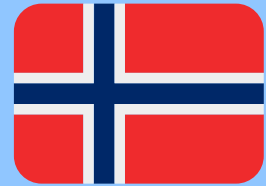
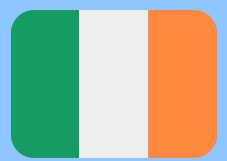
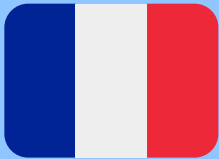
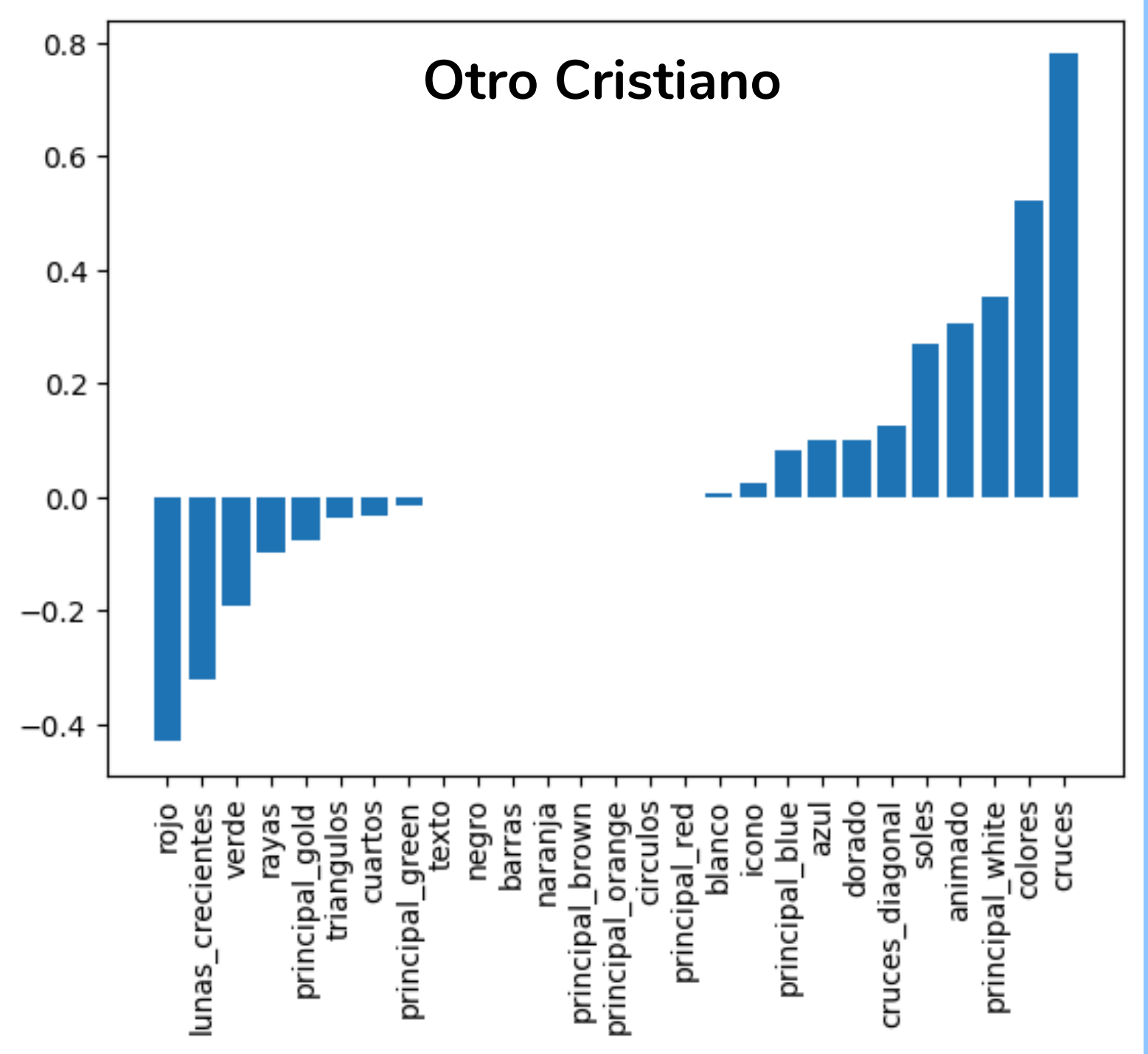
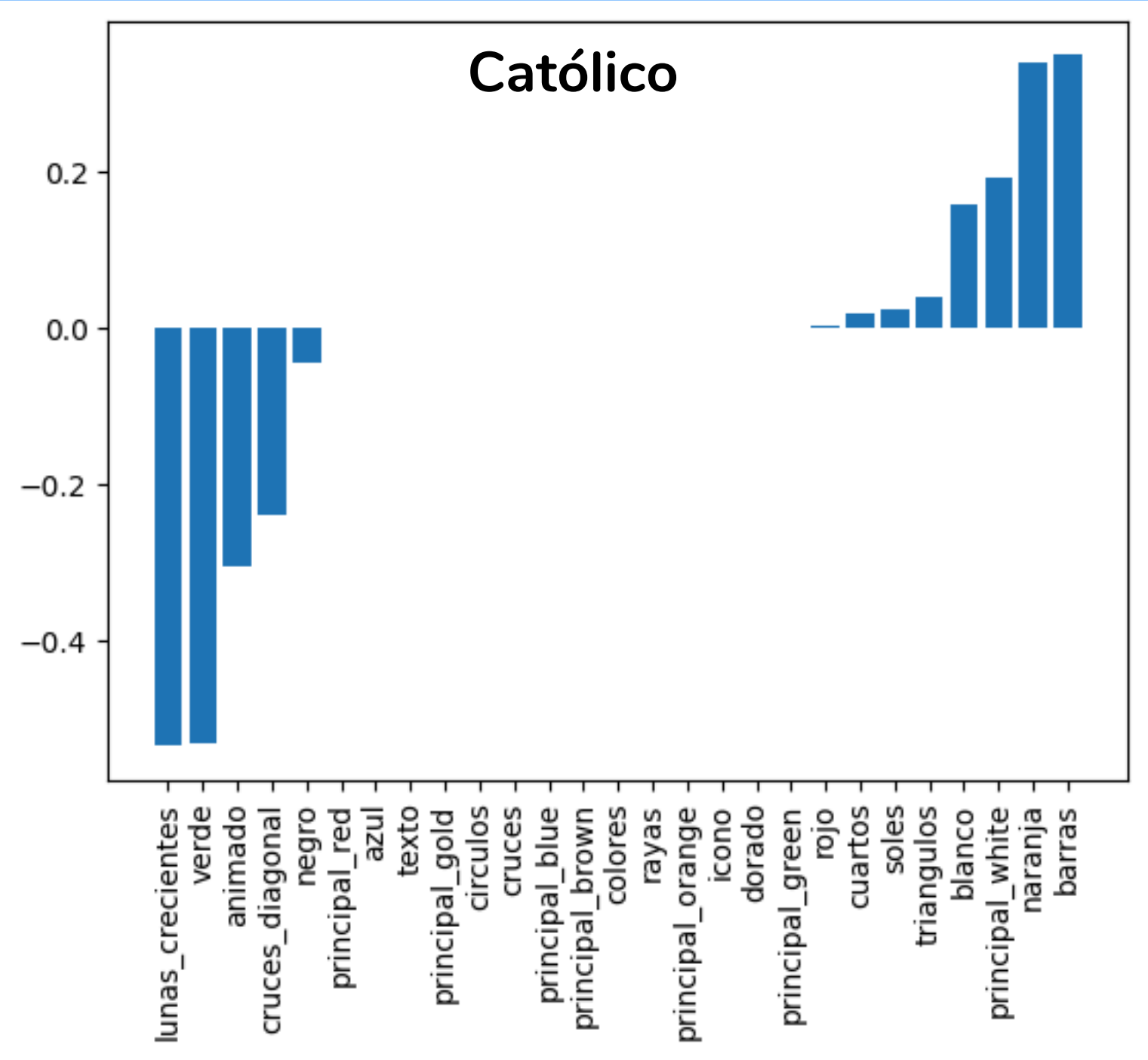
## RELIGIÓN

test\_size=0.20, stratify= y

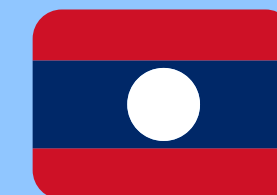
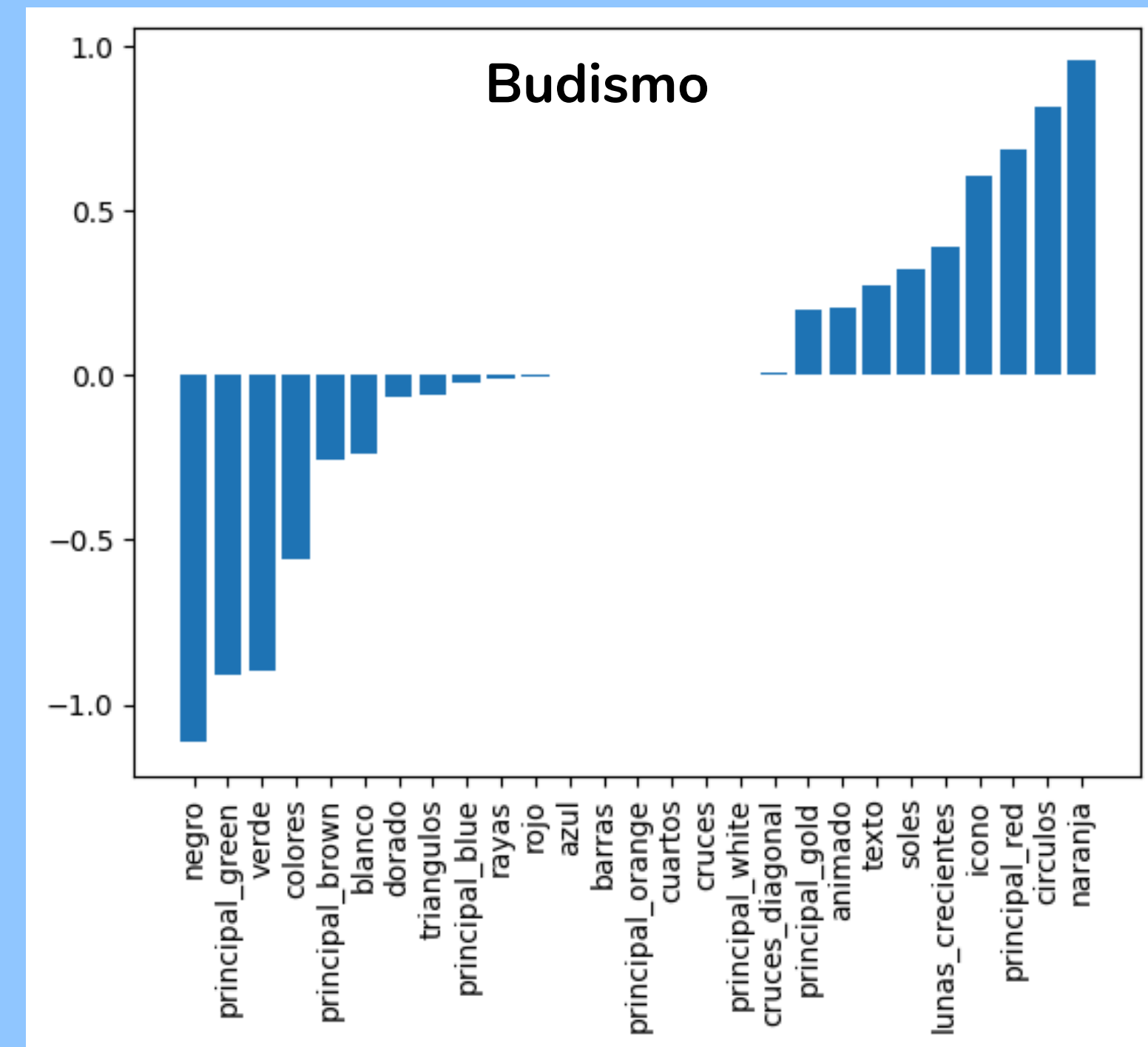
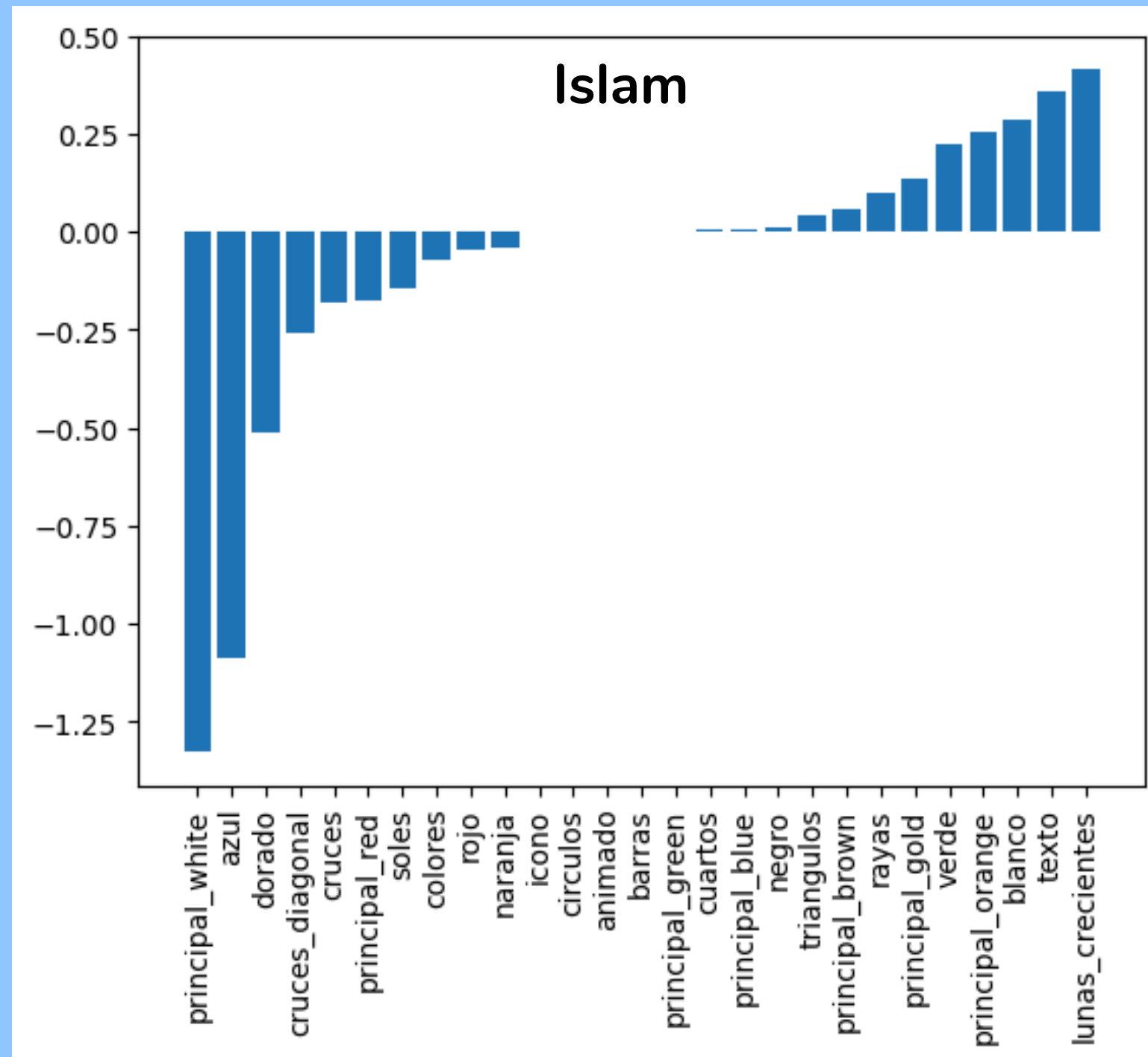
cv=4, scoring='f1\_weighted'

Modelo	Parámetros	F-Score
Random Forest	Default	0.97
Grid Search SVM	C: [ <b>1</b> ,50,100] kernel : [ <b>linear</b> , poly, rbf, sigmoid] degree : [ <b>2</b> ,3,5] gamma: [ <b>scale</b> ,auto]	0.42
Grid Search KNN	n_neighbors: [3,5, <b>10</b> ,20,30] weight : [uniform, <b>distance</b> ] algorithm: [ <b>auto</b> , ball_tree, kd_tree, brute] p: [ <b>1</b> , 2] leaf_size: [ <b>10</b> , 20, 30]	0.49
Grid Search Regresión Logística (Lasso)	C: [0.00001, 0.001, 0.1, <b>0.99</b> , 10, 100, 1000000] class_weight: [dict, <b>balanced</b> ] solver: [liblinear, <b>saga</b> ]	0.47

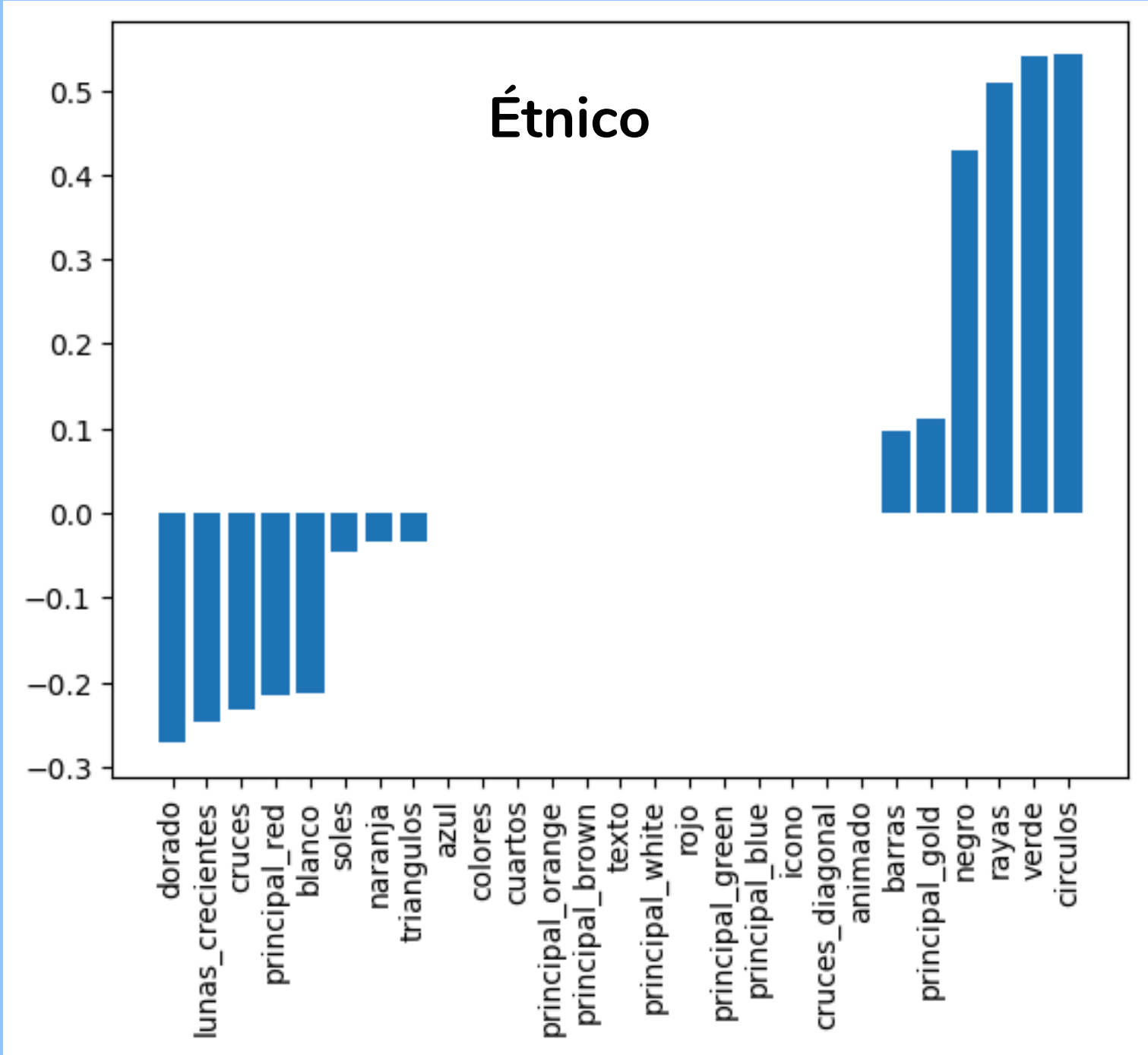
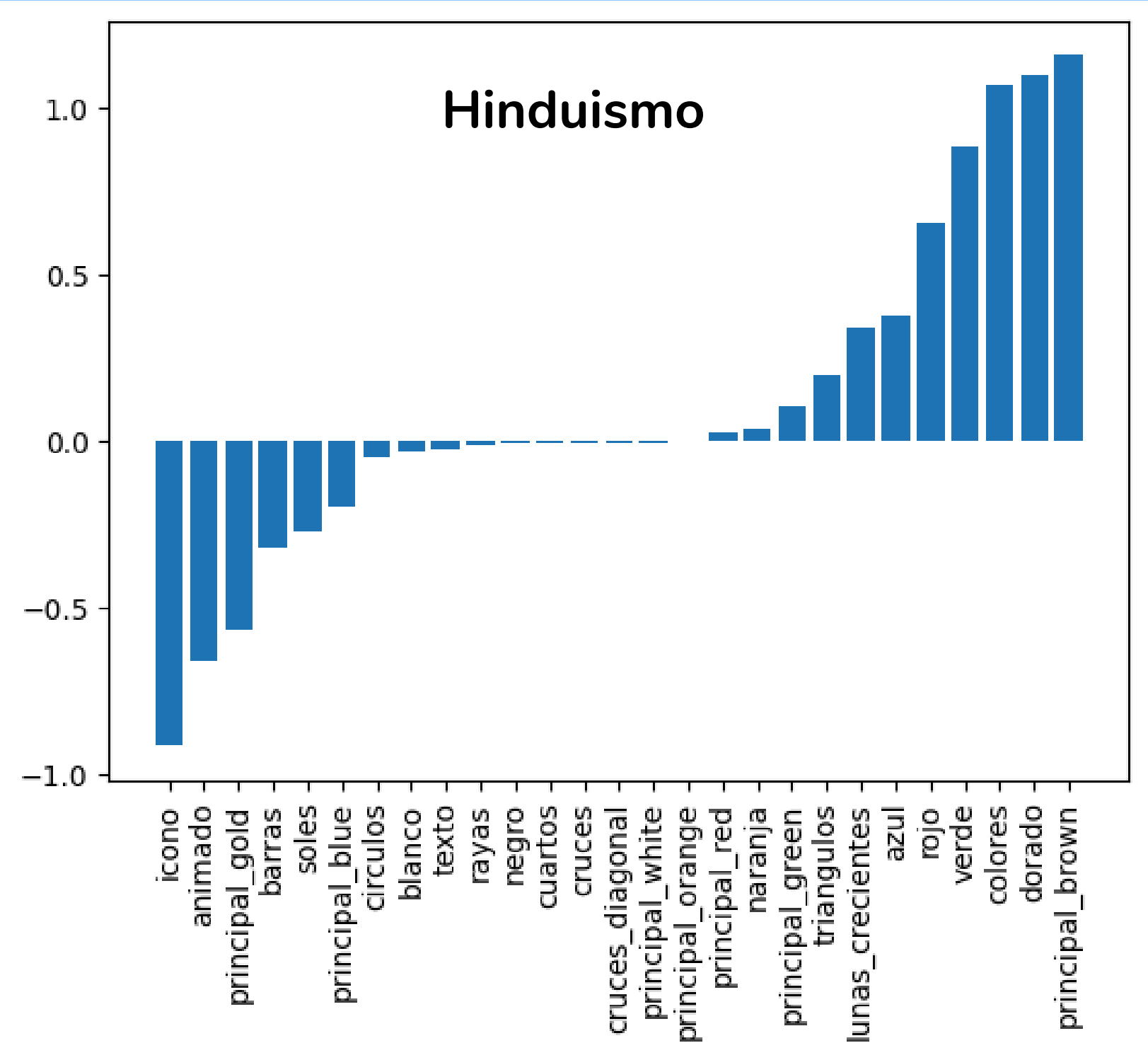
# Reg. Log. - Selección de Features



# Reg. Log. - Selección de Features

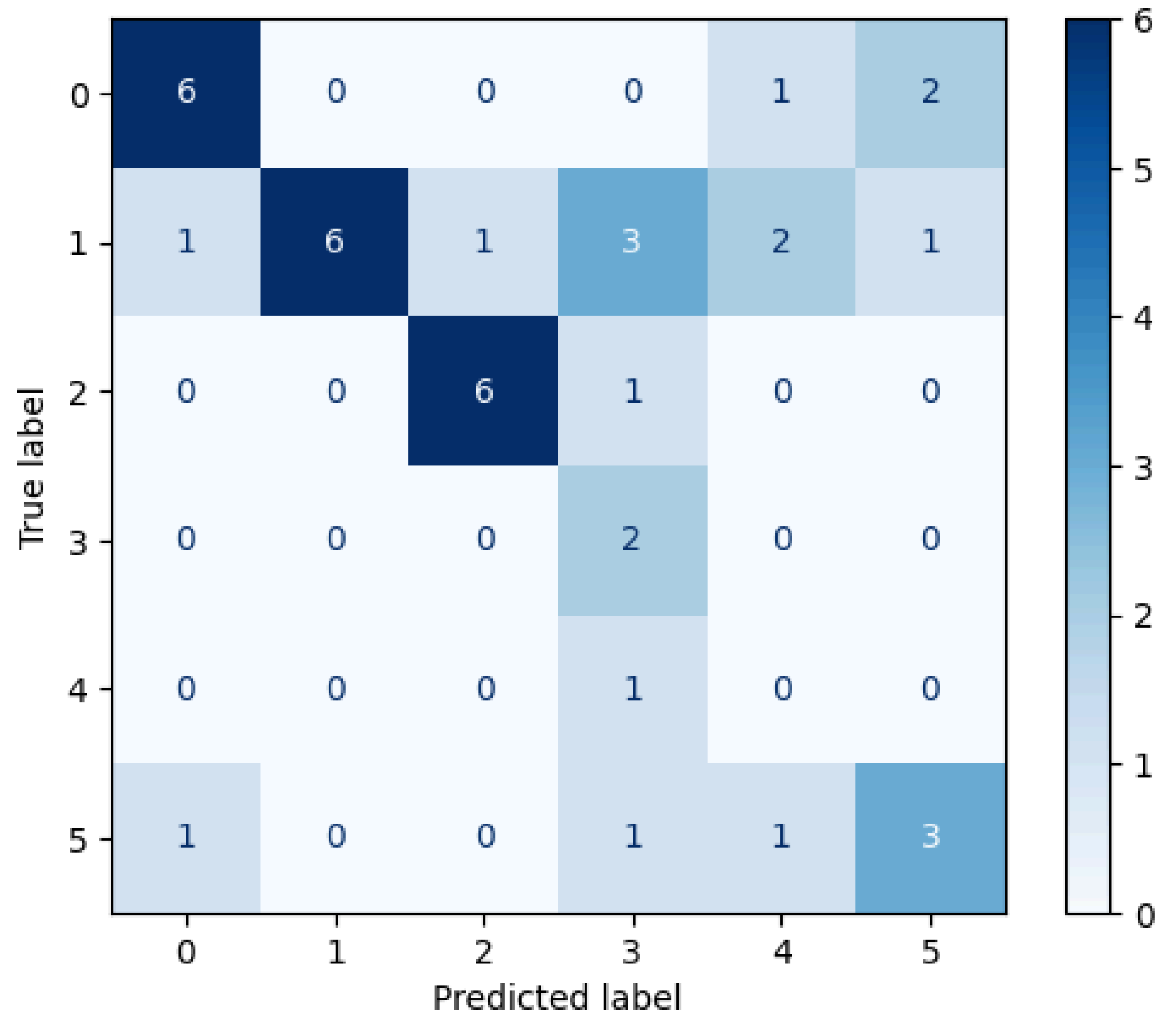


# Reg. Log. - Selección de Features








# Evaluación Religión

Clase	Precisión	Recall	F1-score
0 = Católico	0.75	0.67	0.71
1 = Otro cristiano	1.00	0.43	0.60
2 = Musulmán	0.86	0.86	0.86
3 = Budista	0.25	1.00	0.40
4 = Hindú	0.00	0.00	0.00
5 = Étnico	0.50	0.50	0.50



# Evaluación

	Aciertos	Religión
	Argentina	Católico
	Dinamarca	Otro Cristiano
	Emiratos Árabes Unidos	Musulmán
	Gabón	Étnico
	Buán	Budismo

	Errores	Religión real	Predicción
	Singapur	Budista	Musulmán
	Ruanda	Étnico	Musulmán
	USSR	Otro cristiano	Budista
	Portugal	Católico	Otro cristiano
	Bolivia	Católico	Étnico

# Métricas y modelos

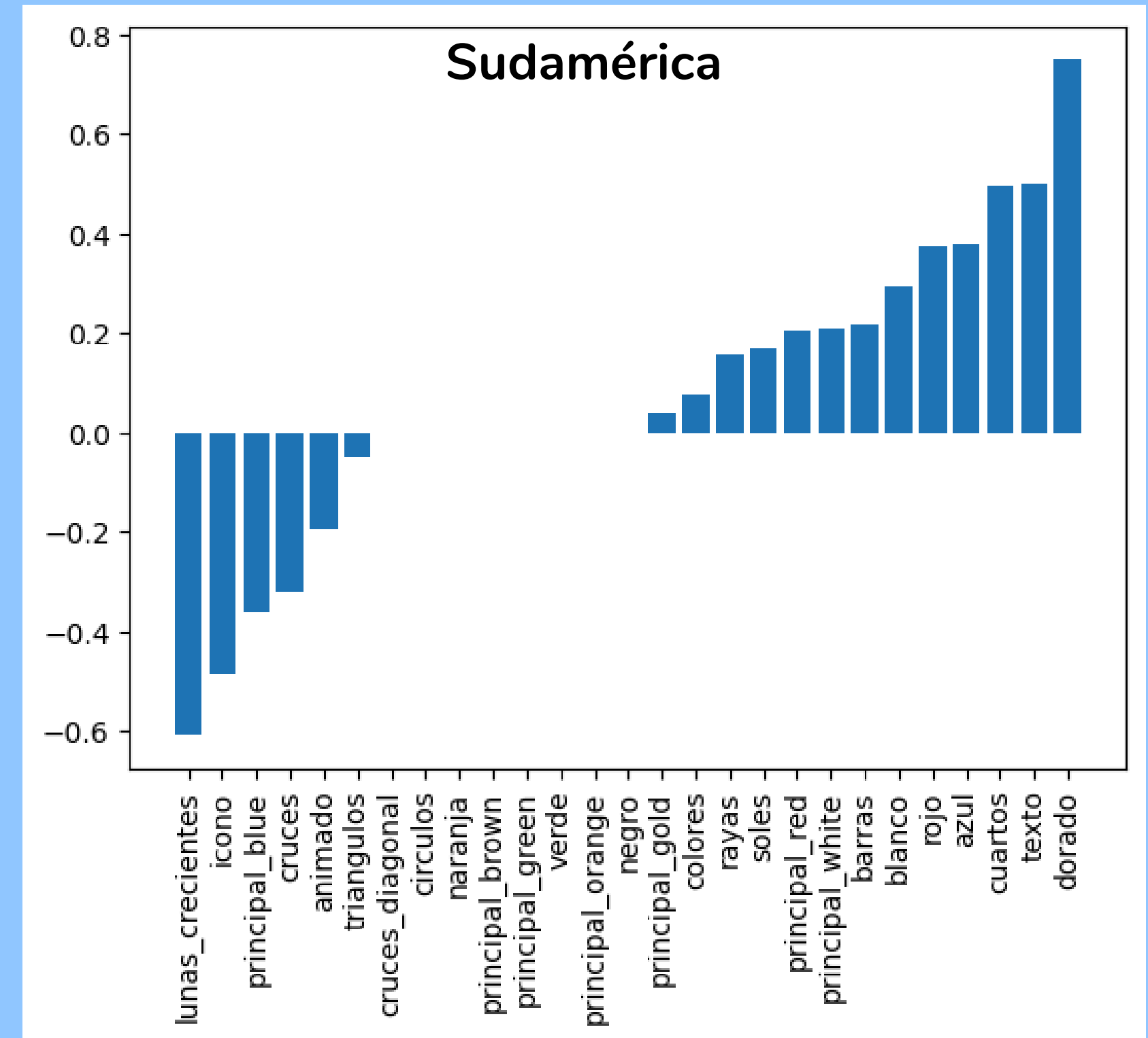
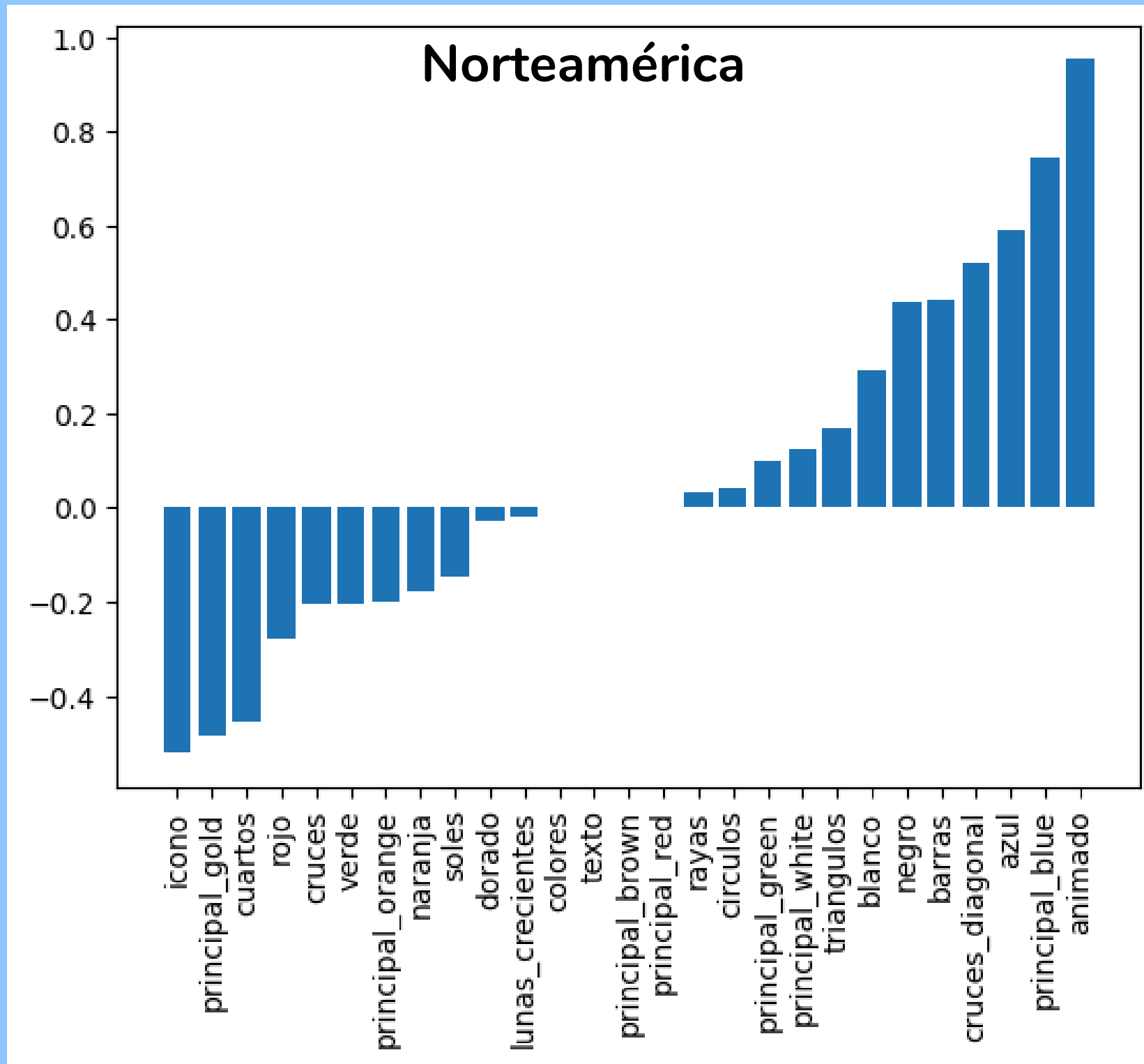
## CONTINENTE

test\_size=0.20, stratify= y

cv=4, scoring='f1\_weighted'

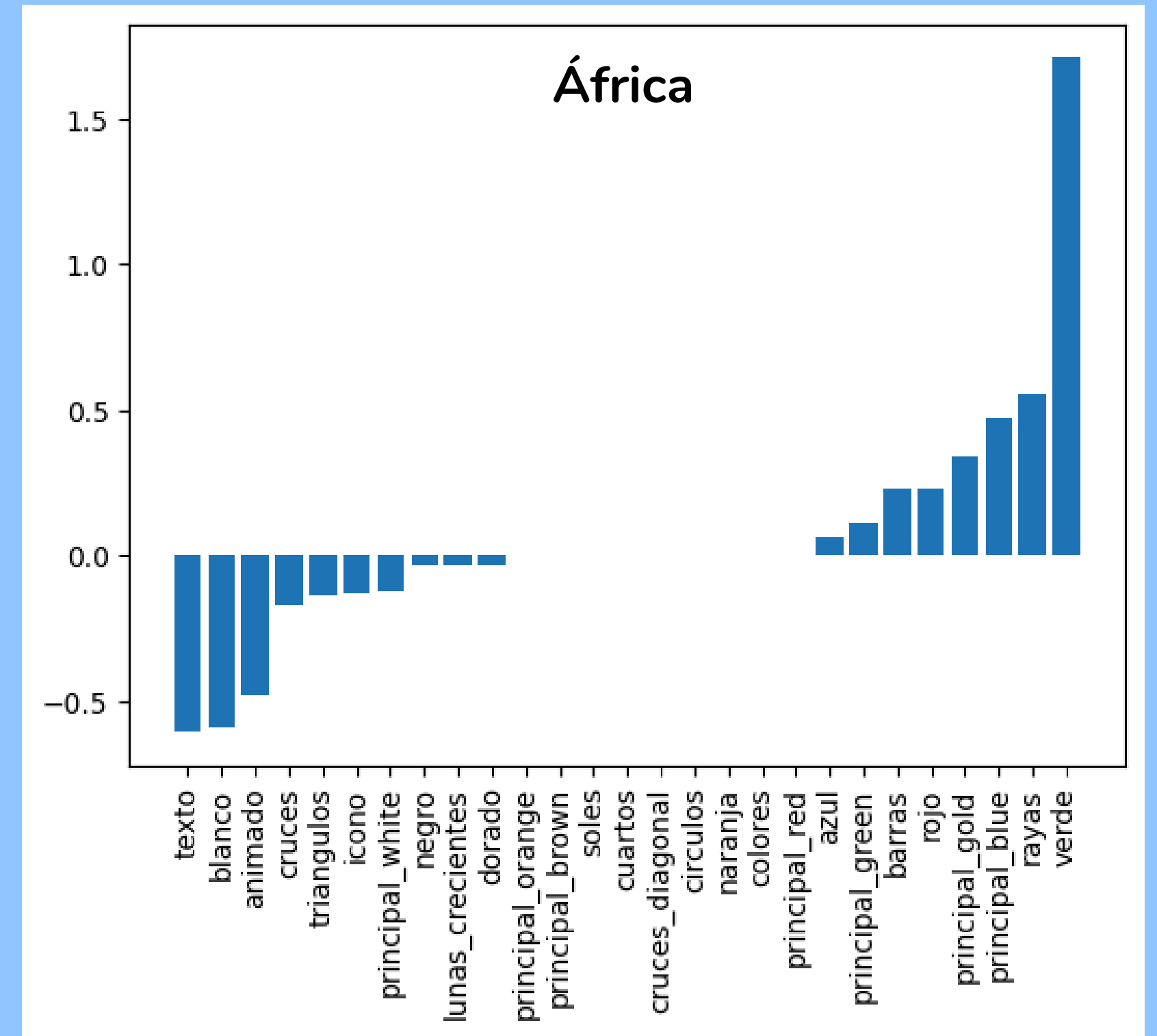
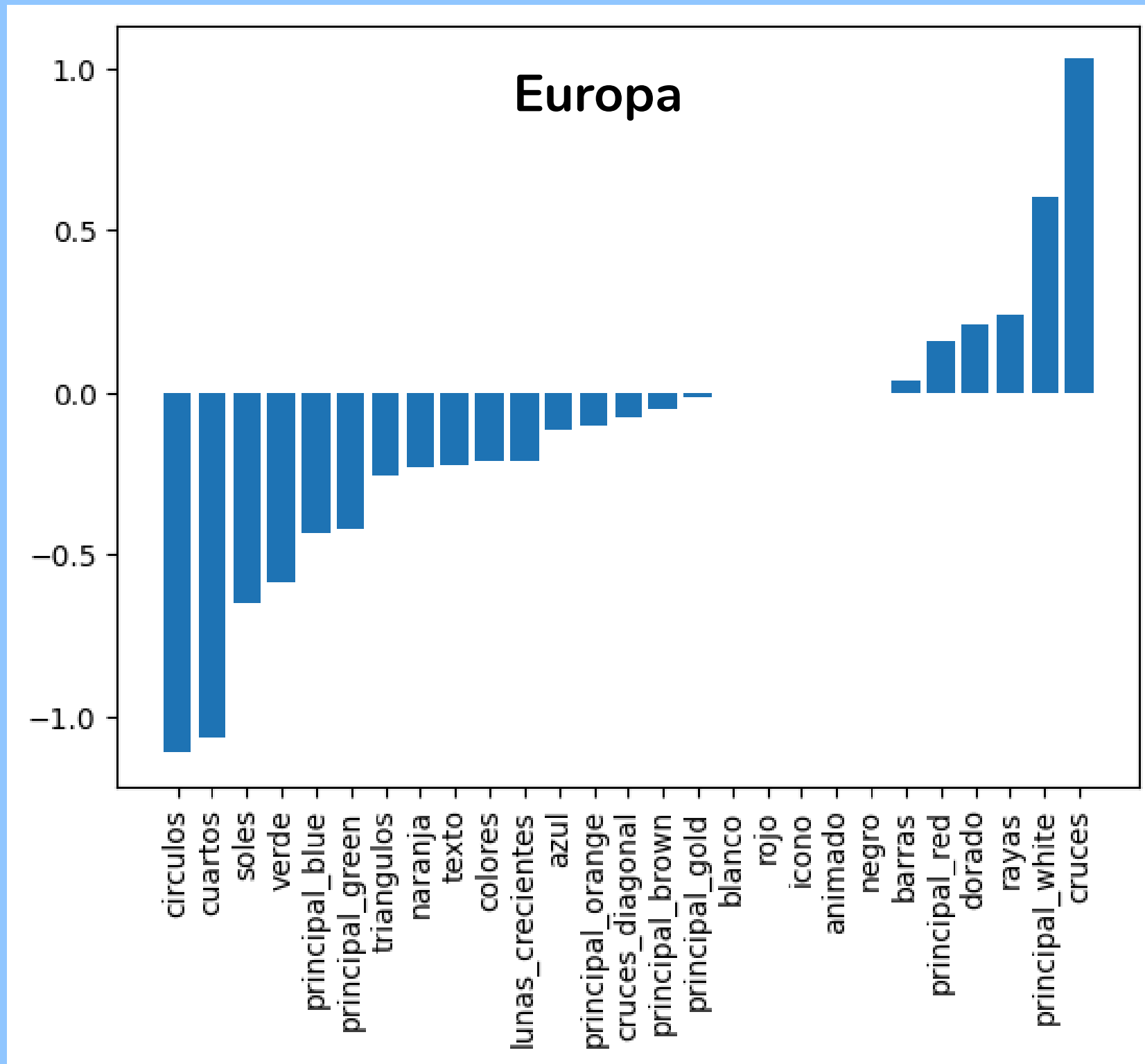
Modelo	Parámetros	F-Score
Grid Search Regresión Logística (Lasso)	C: [0.00001, 0.001, 0.1, <b>0.99</b> , 10, 100, 1000000] class_weight: [dict, <b>balanced</b> ] solver: [liblinear, <b>saga</b> ]	0.41

# Continente - Selección de Features

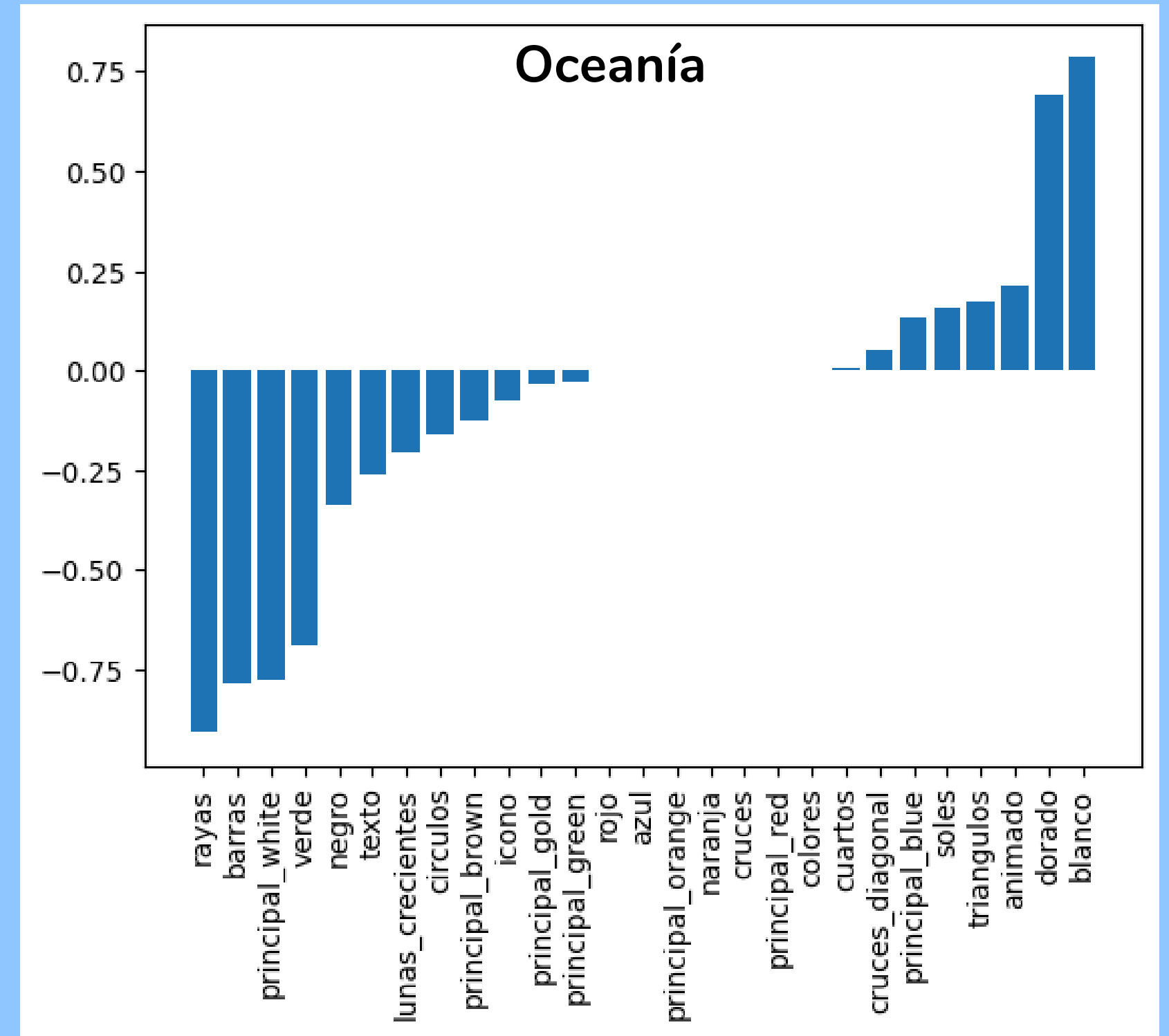
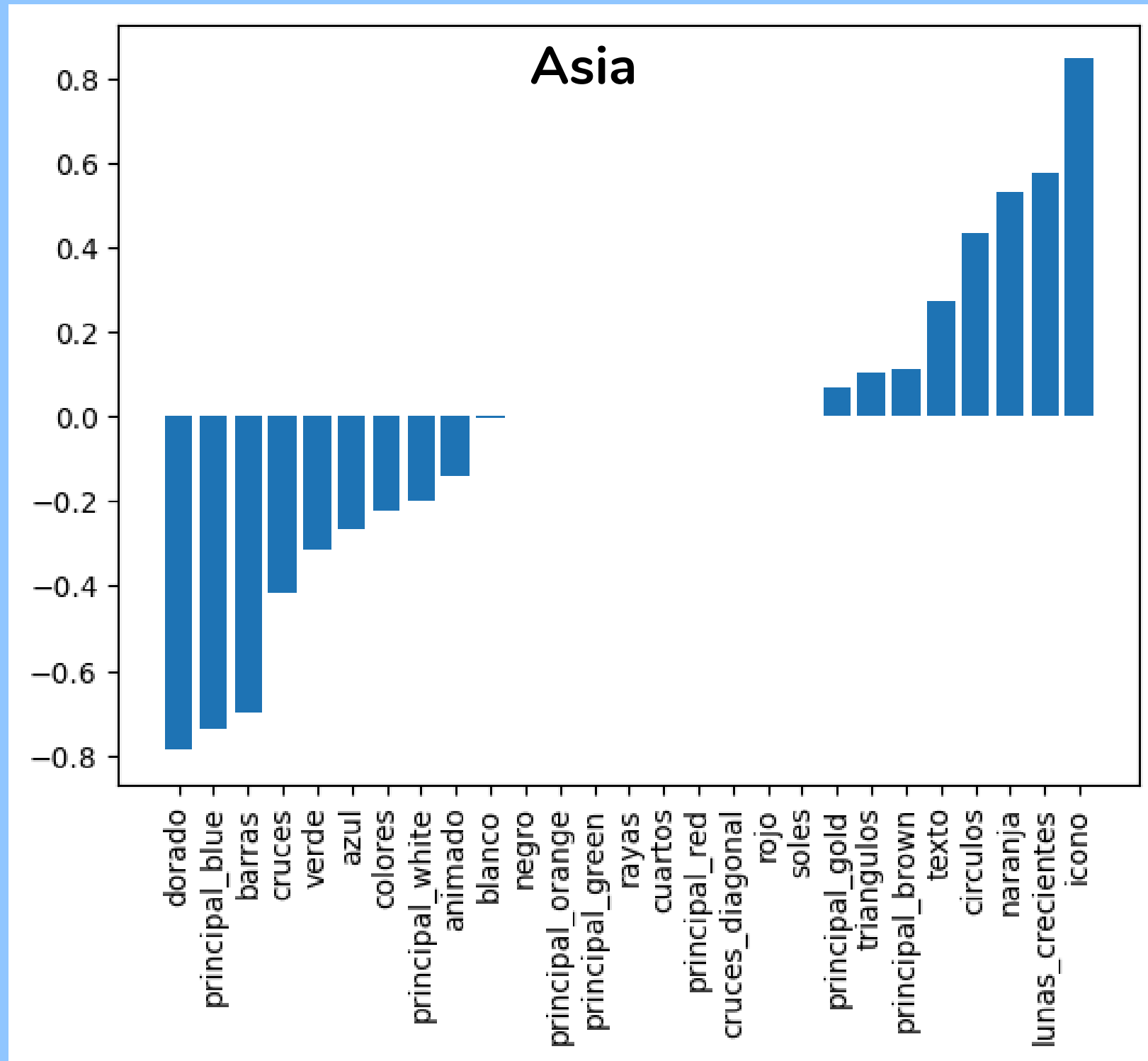




# Continente - Selección de Features

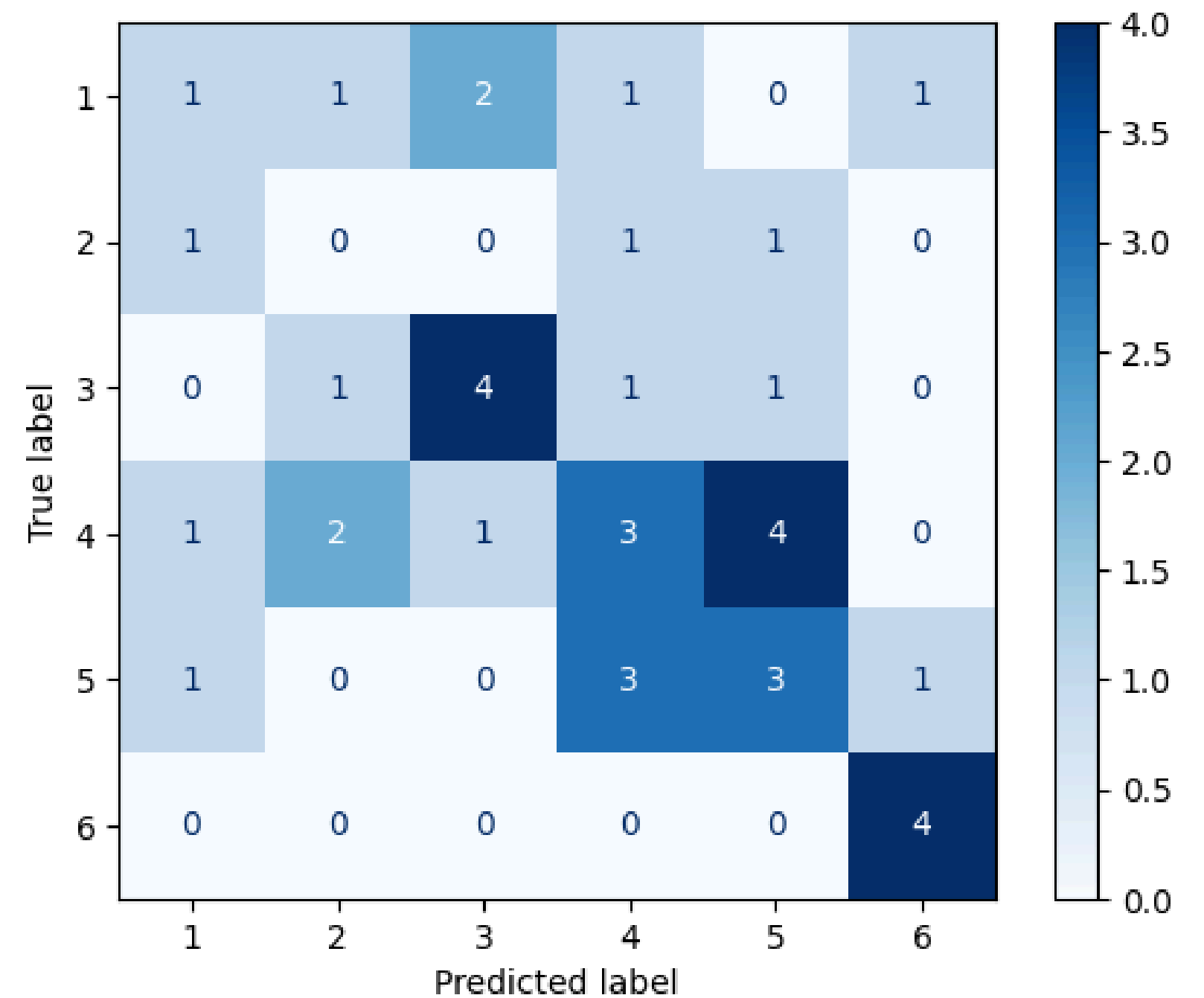


# Continente - Selección de Features



# Evaluación Continente

Clase	Precisión	Recall	F1-score
0 = NAmerica	0.25	0.17	0.20
1 = SAmerica	0.00	0.00	0.00
2 = Europa	0.57	0.57	0.57
3 = Africa	0.33	0.27	0.30
4 = Asia	0.33	0.38	0.35
5 = Oceania	0.67	1.00	0.80



# Métricas y modelos

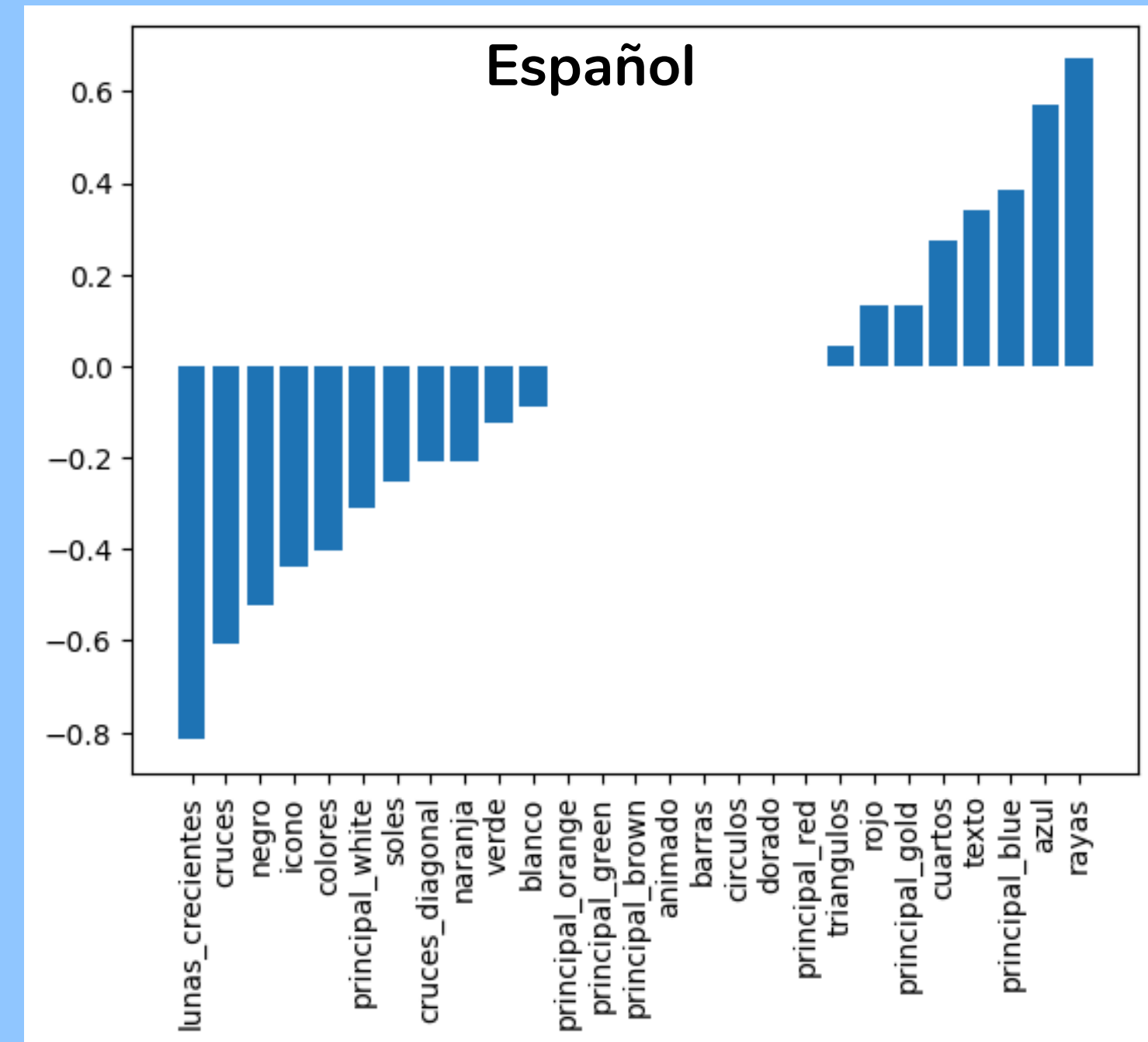
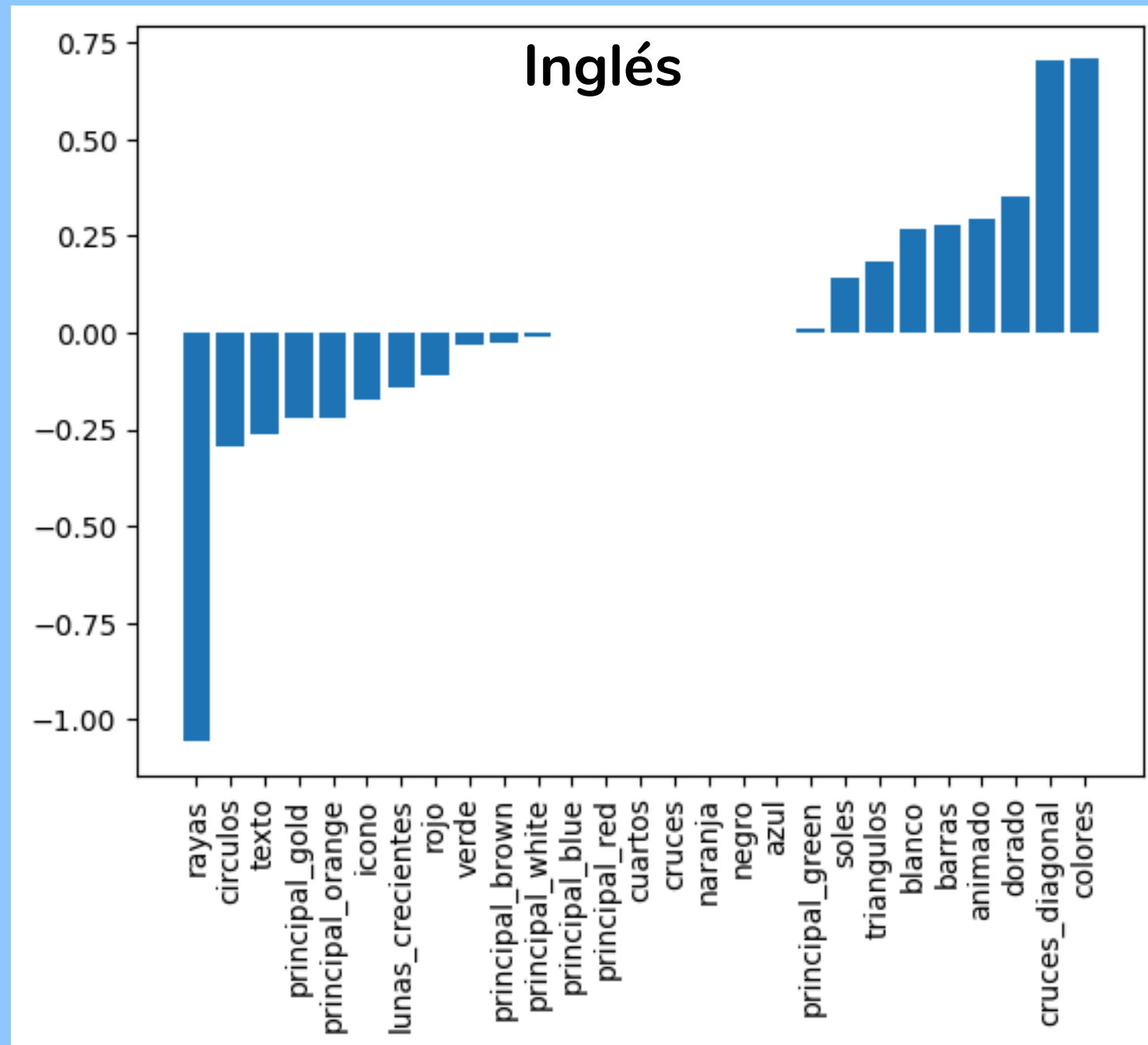
## IDIOMA

test\_size=0.20, stratify= y

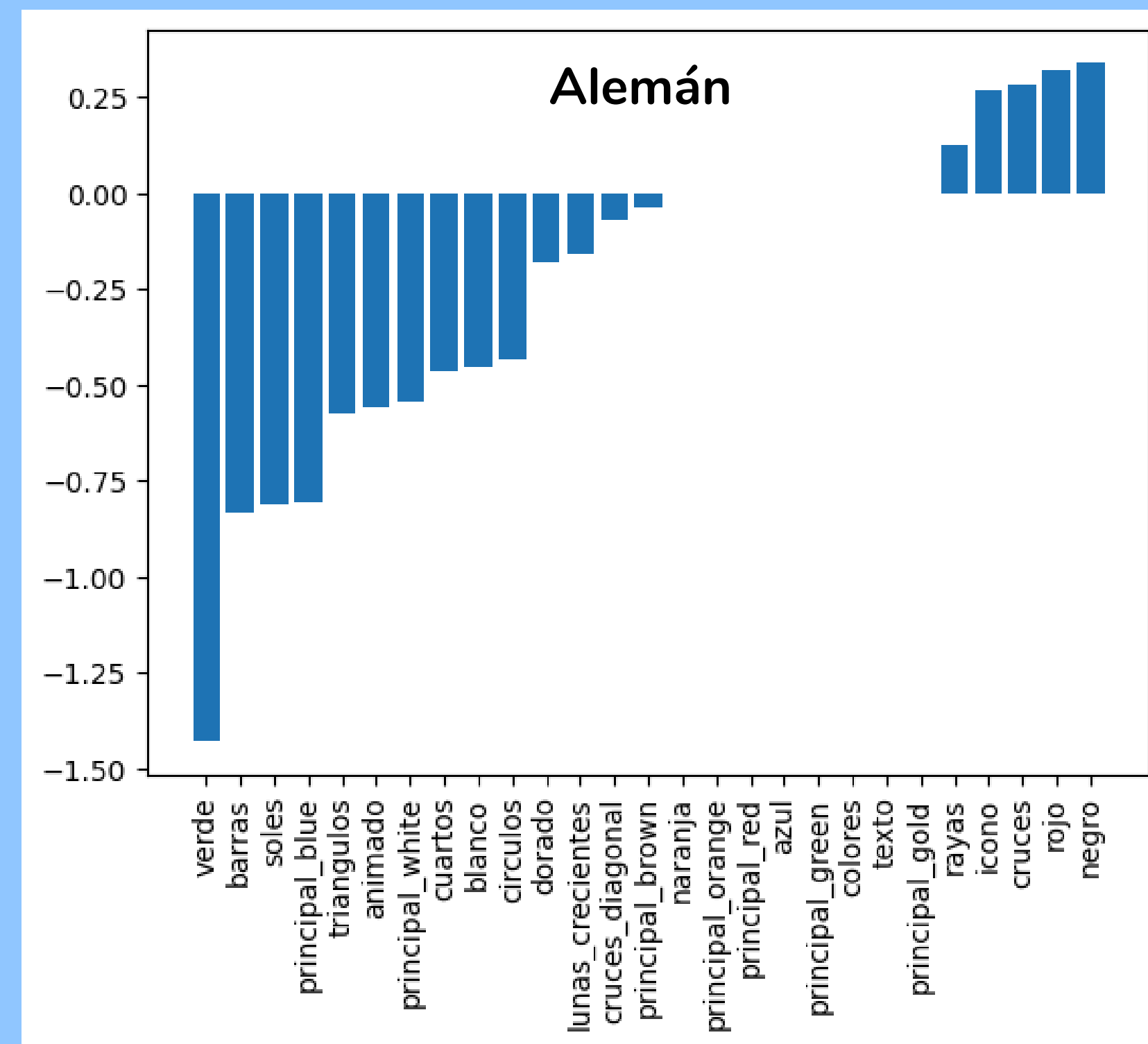
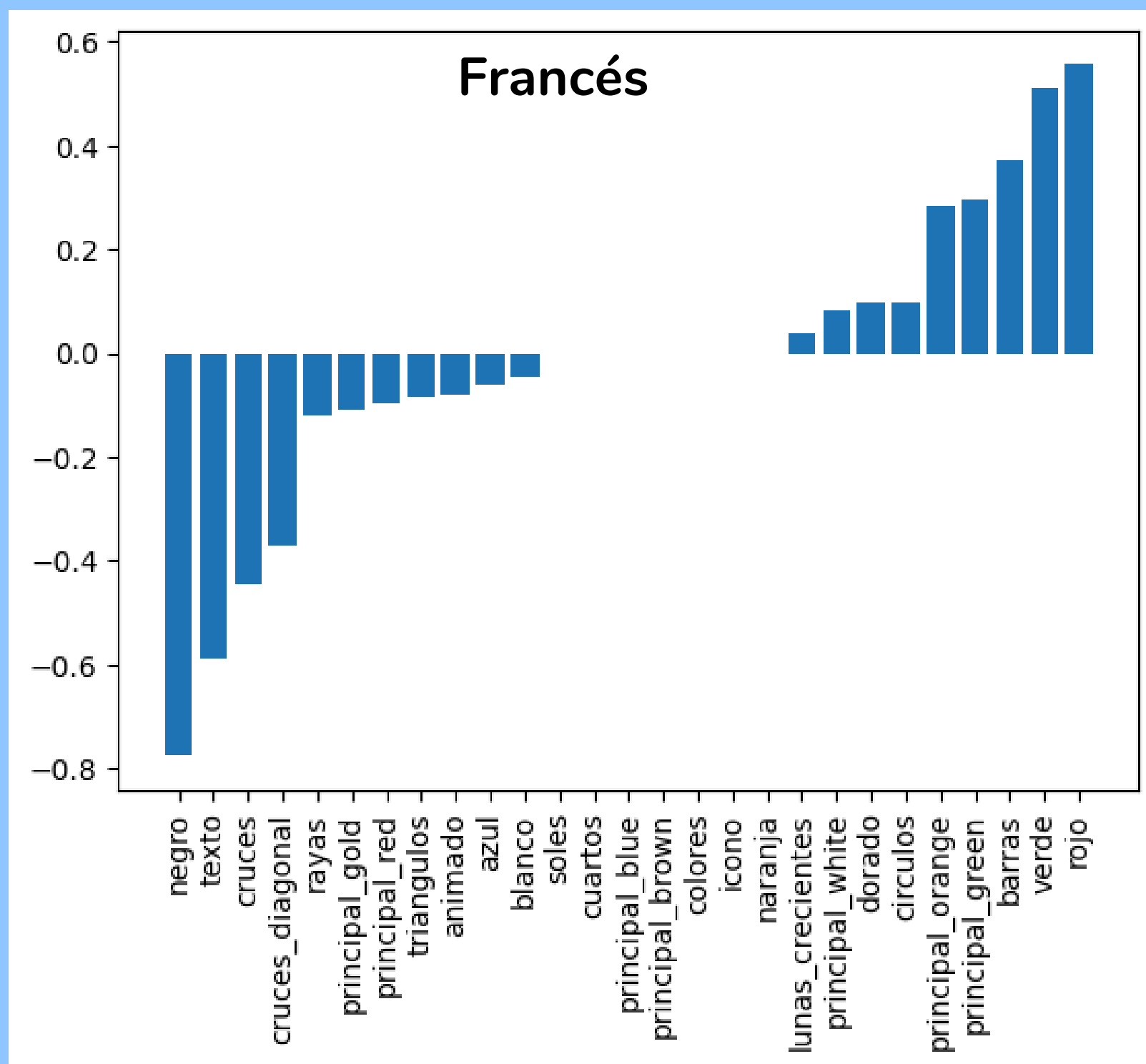
cv=4, scoring='f1\_weighted'

Modelo	Parámetros	F-Score
Grid Search Regresión Logística (Lasso)	C: [0.00001, 0.001, 0.1, <b>0.99</b> , 10, 100, 1000000] class_weight: [dict, <b>balanced</b> ] solver: [liblinear, <b>saga</b> ]	0.41

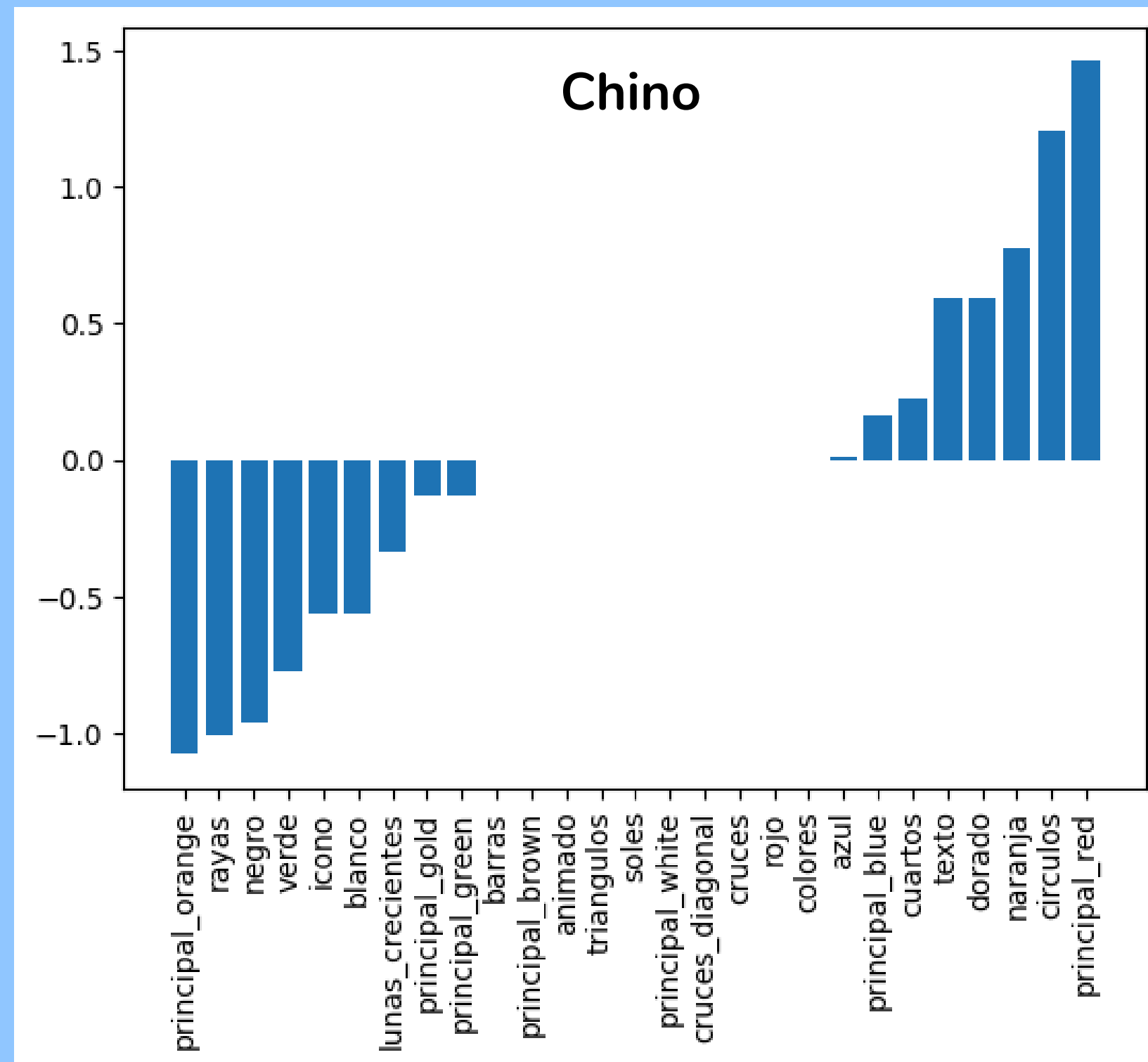
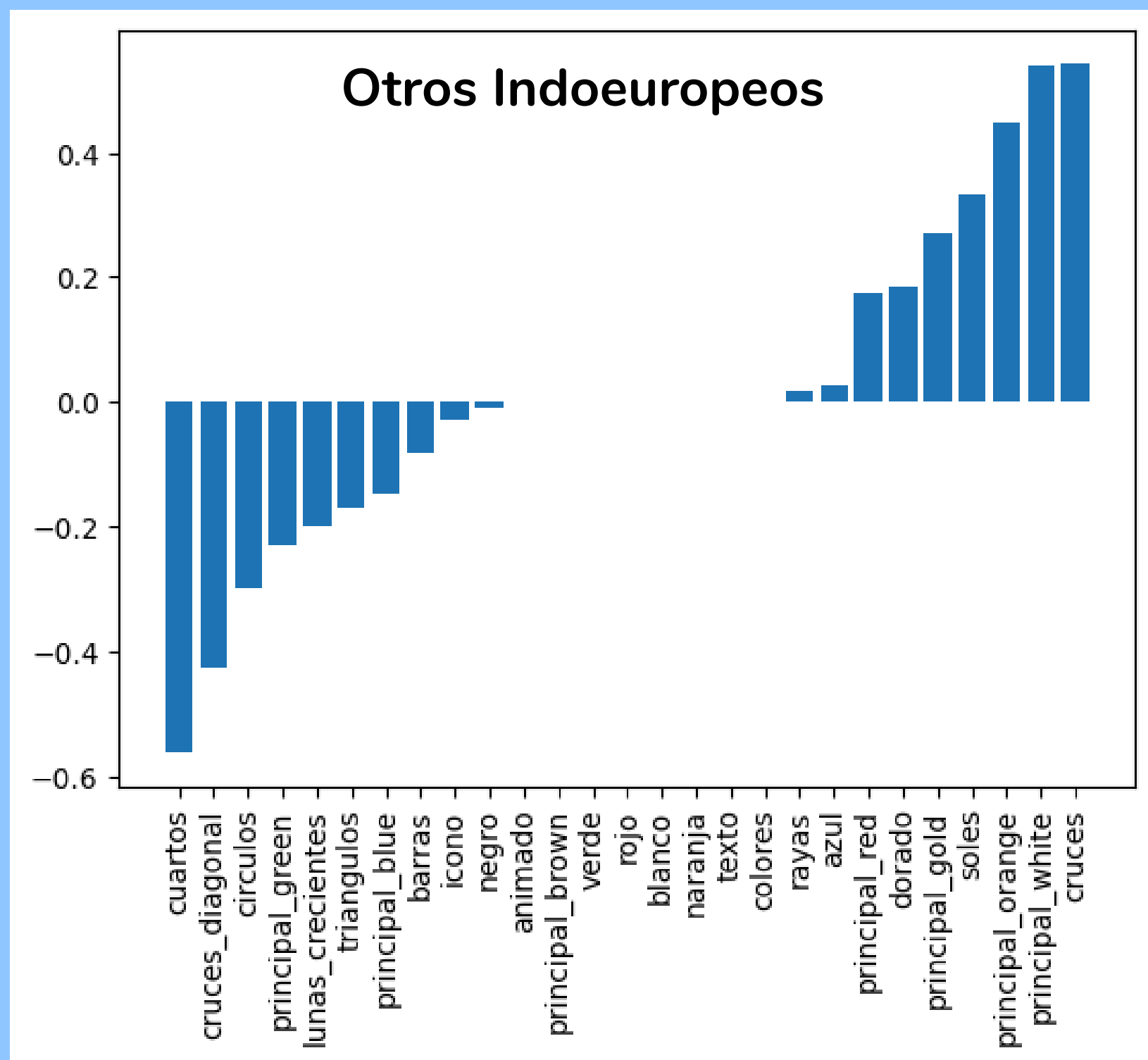
# Idioma - Selección de Features



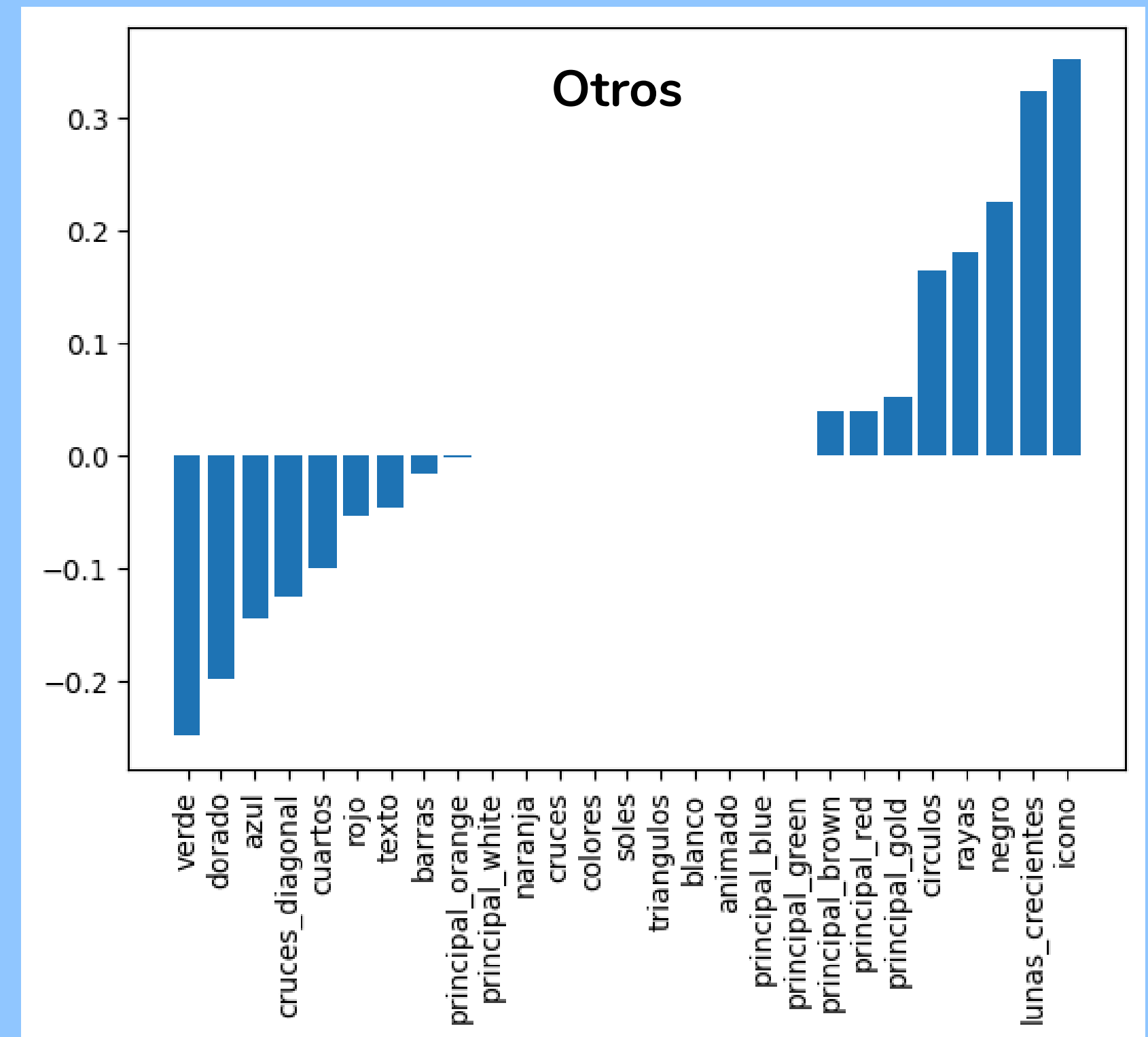
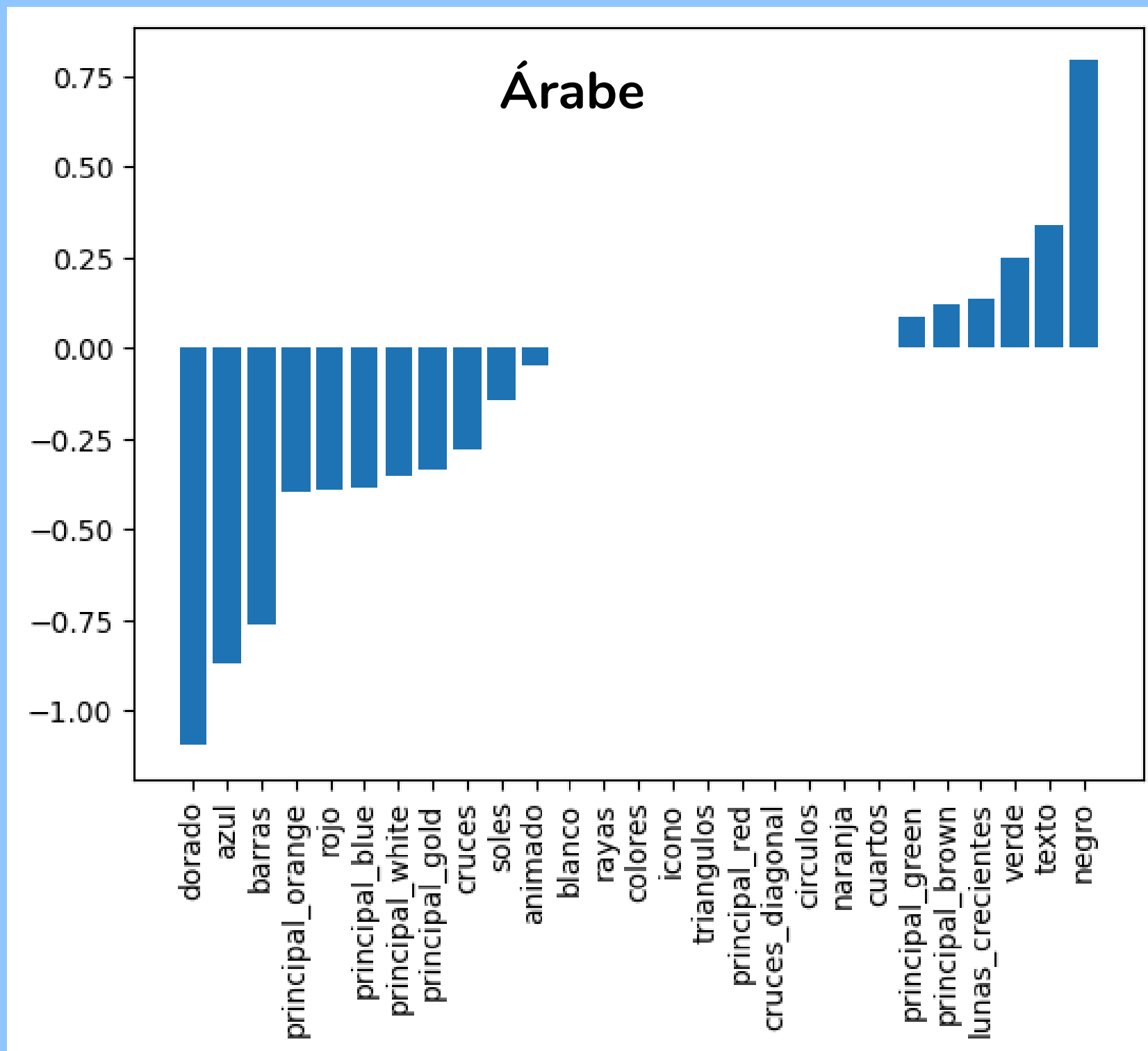
# Idioma - Selección de Features



# Idioma - Selección de Features



# Idioma - Selección de Features

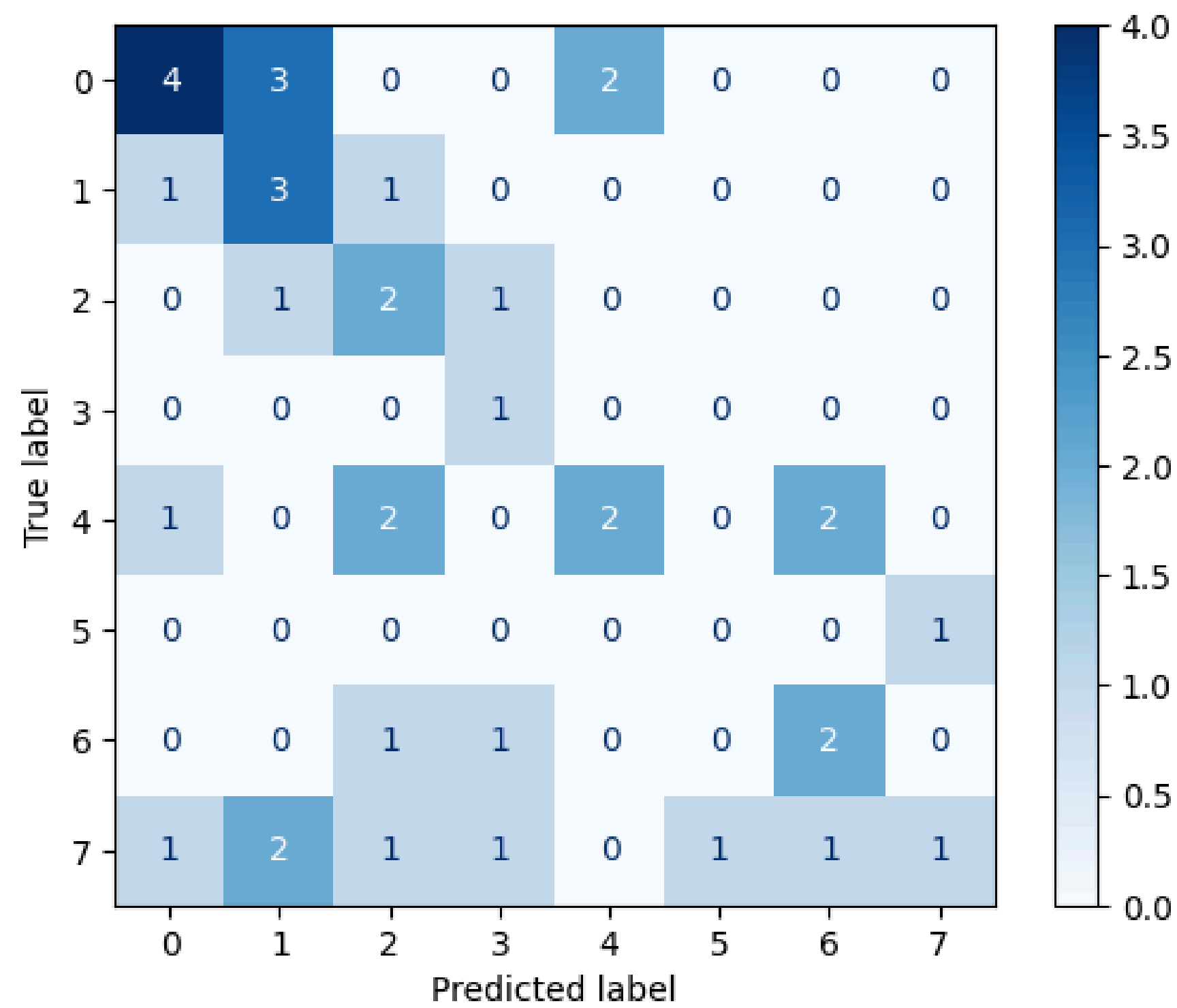




# Evaluación

# Idioma

Clase	Precisión	Recall	F1-score
0 = inglés	0.57	0.44	0.50
1 =Español	0.33	0.60	0.43
2 = Francés	0.29	0.50	0.36
3 =Alemán	0.25	1.00	0.40
4 = Otros Indo-Europeos	0.50	0.29	0.36
5=Chino	0.00	0.00	0.00
6= Árabe	0.40	0.50	0.44
7 =Otros	0.50	0.12	0.20



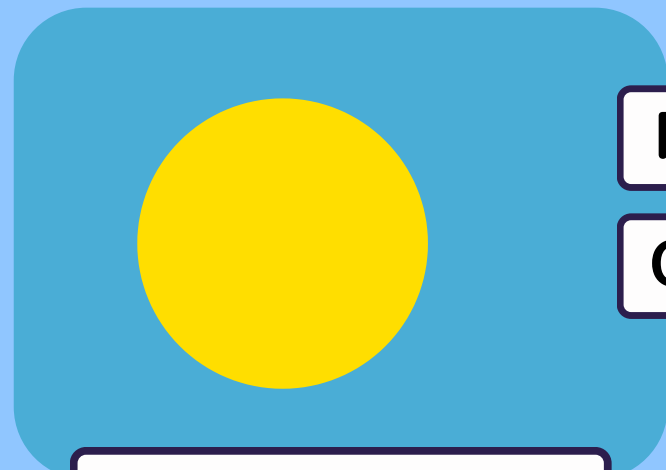
# Desafiando el modelo



Eritrea

Religión ✗

Continente ✓



Palaos

Religión ✗

Continente ✗



Turkmenistan

Religión ✗

Continente ✗



Sudán del sur

Religión ✗

Continente ✗



Georgia

Religión ✗

Continente ✓



Montenegro

Religión ✗

Continente ✓



Serbia

Religión ✗

Continente ✗



Kazajistan

Religión ✗

Continente ✗

# Conclusiones

En definitiva, se puede concluir que las banderas representan en alguna medida, más allá de lo simbólico, aspectos culturales, lingüísticos y geográficos de cada país. A través de su diseño en ciertos casos podríamos llegar a estimar la religión, el idioma y el continente.

## Posibles mejoras

- 🚩 Uso de datos oficiales
- 🚩 Países actuales
- 🚩 Permitir que por país se registre mas de una religión e idioma

## Consideraciones

- 🚩 Paso del tiempo



# iGracias!

¿Alguna pregunta?