MENTORIA

Clasificación de Tumoresferas

GRUPO 1 - 2023

PRESENTACIÓN FINAL

ZULATTO, MARIA LAURA
APARICIO,ERIKA SABRINA
POZO ARCE, IGNACIO
CIFUENTES SAMBONI, UBEIDEN

OBJETIVO GENERAL

Analizar los datos obtenidos por procesamiento de imágenes de cultivos de células tumorales y de estructuras celulares denominadas TUMORESFERAS.

Desarrollar un modelo computacional que permita clasificar las mismas como tal y así apoyar en la toma de decisión final del personal calificado.

ALCANCE Y CONTENIDO

1. BREVE RESUMEN PREVIO
2. EXPLORACIÓN DE DATOS
3. MODELOS DE APRENDIZAJE
4. CONCLUSIONES Y OBSERVACIONES

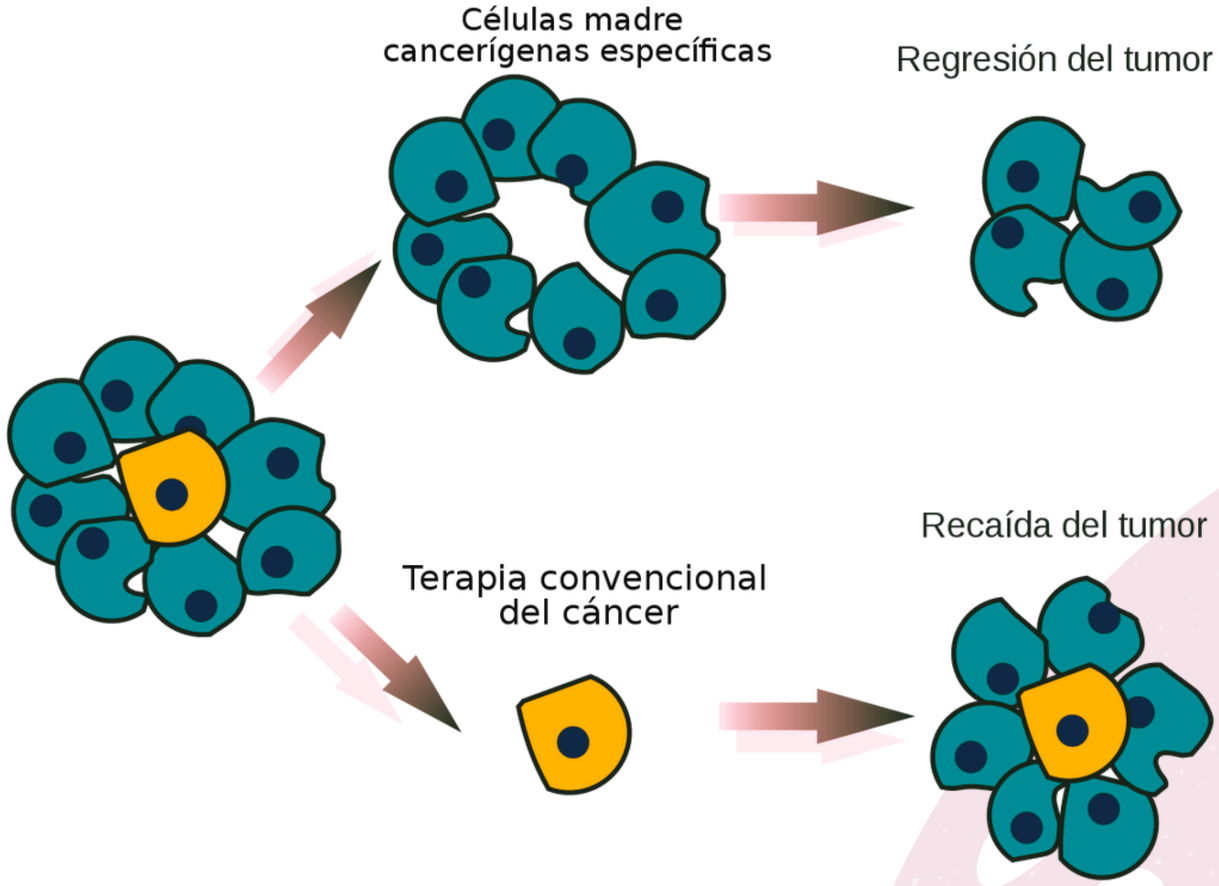
RECORDAMOS!

¿Qué es una Tumoresfera?

Una Tumoresfera es una estructura celular esferoidal originada a partir de una única célula madre cancerosa (CMC) y que crece en suspensión en cultivos celulares in vitro de células cancerosas.

¿Cuál es la importancia de su estudio?

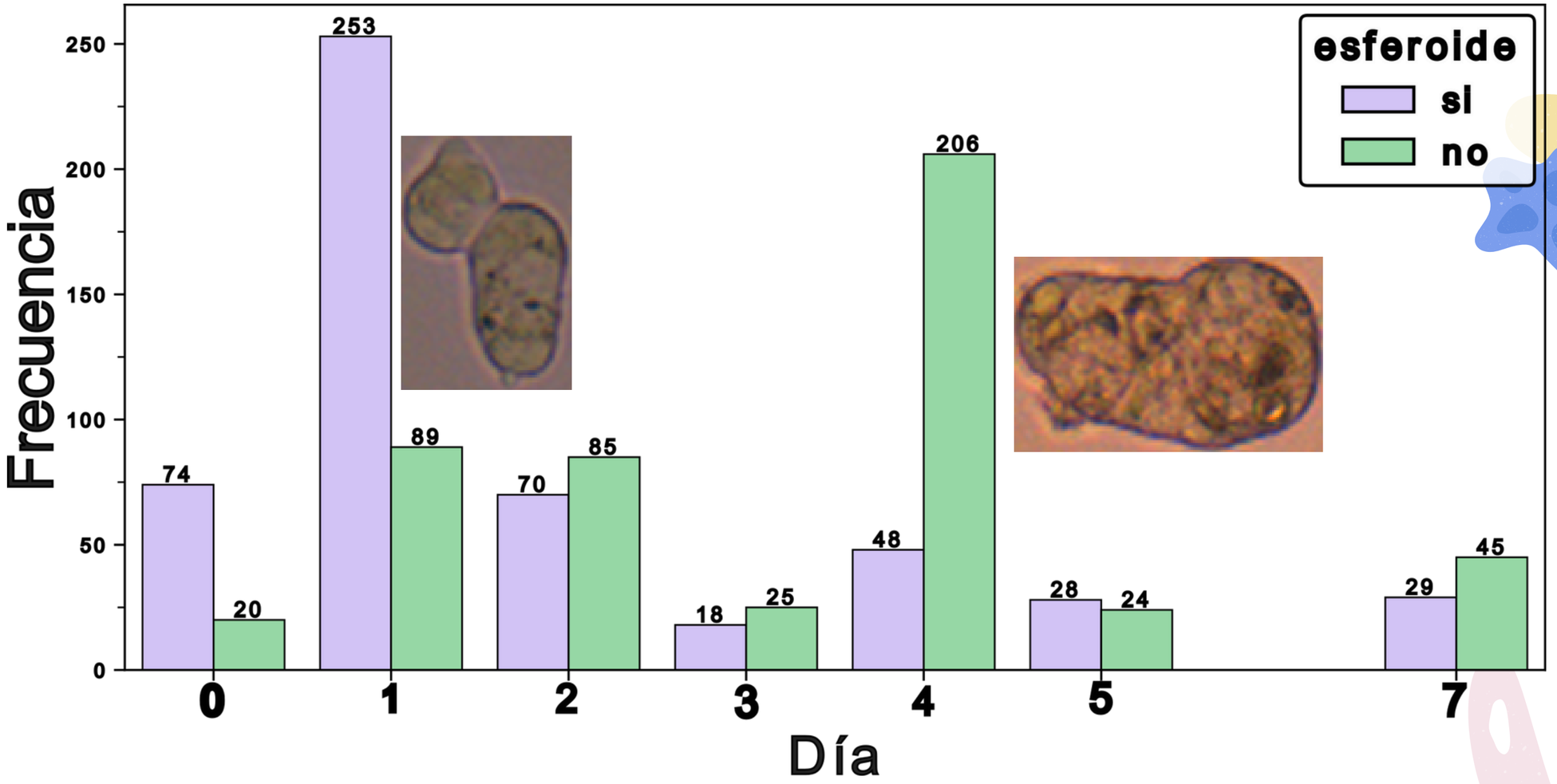
Analizar las tumoresferas podría ayudar a plantear mejoras o alternativas para que los tratamientos anticancerígenos sean mas eficientes.

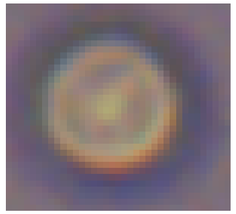


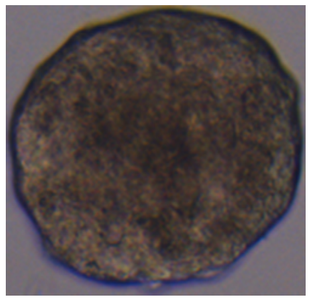
Alcance de 1° práctico

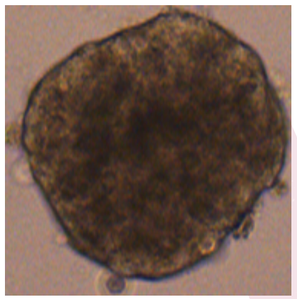
- INTRODUCCIÓN BASE DE DATOS
- ANÁLISIS EXPLORATORIO Y CURACIÓN
- ANÁLISIS DE CORRELACIÓN
- EXPLORACIÓN DE MUESTRAS POR DÍA
- CONCLUSIONES Y OBSERVACIONES

EXPLORACIÓN DE DATOS

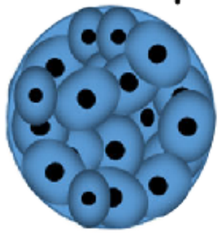

esferoide
sí
no
Frecuencia
Día
0
50
100
150
200
250
0
1
2
3
4
5
7
74
20
253
89
70
85
18
25
48
206
28
24
29
45



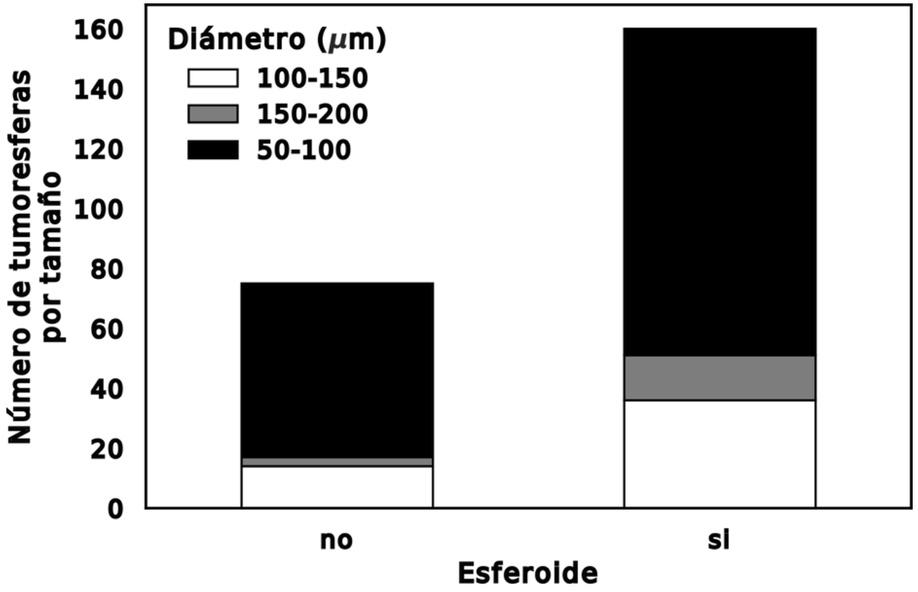


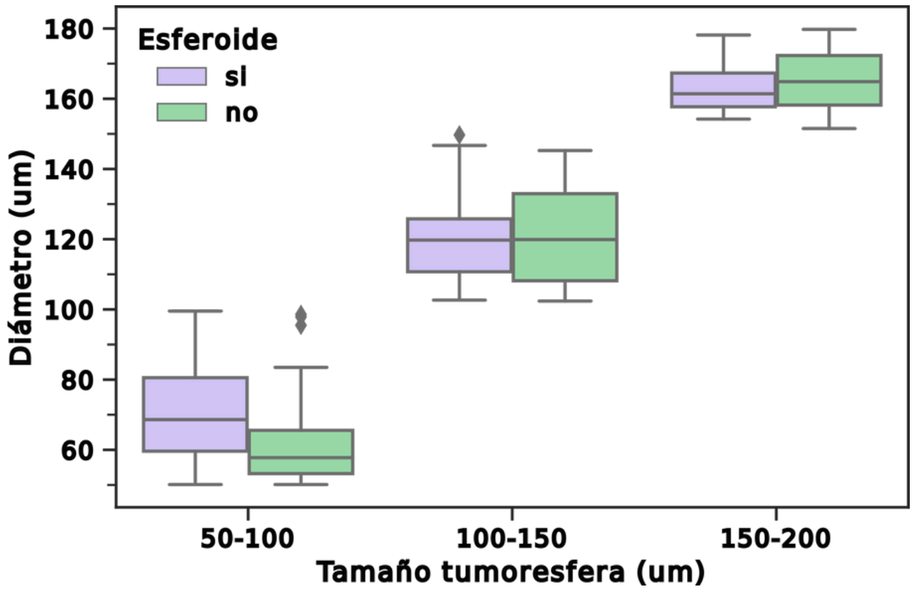


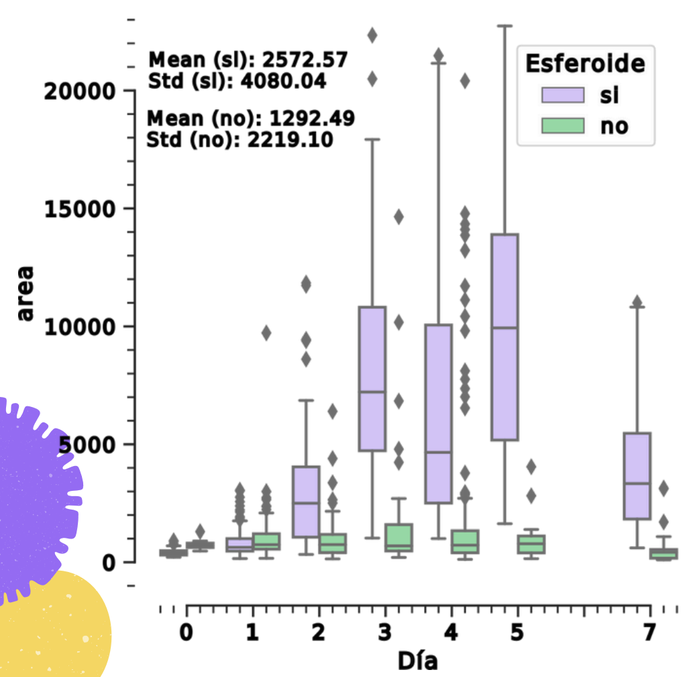
circularidad >0.3
A>100 µm²

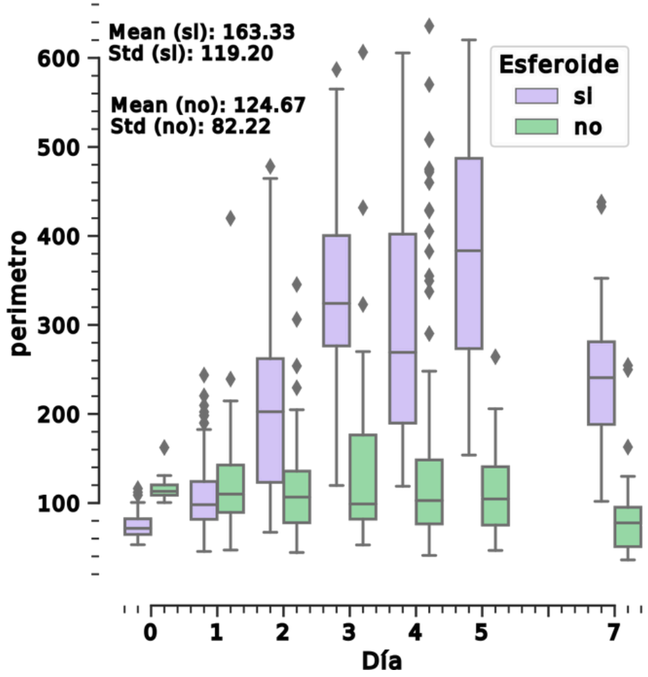


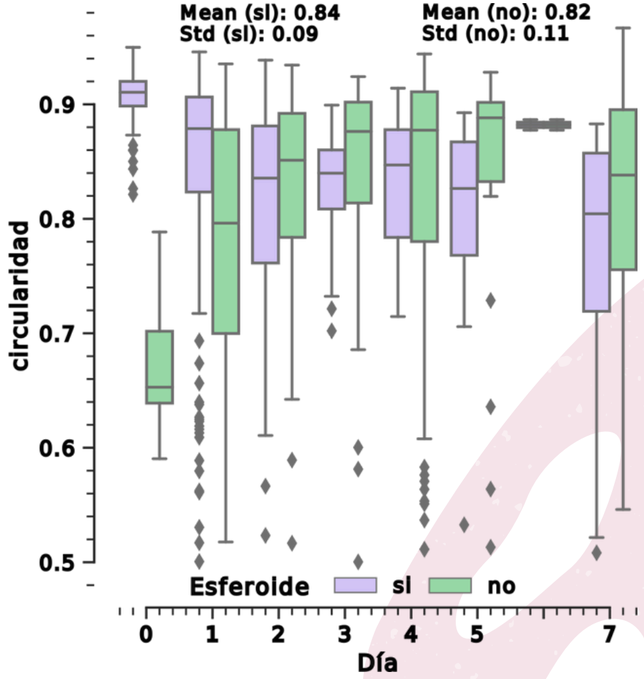
50-250 µm











Pocos datos!!

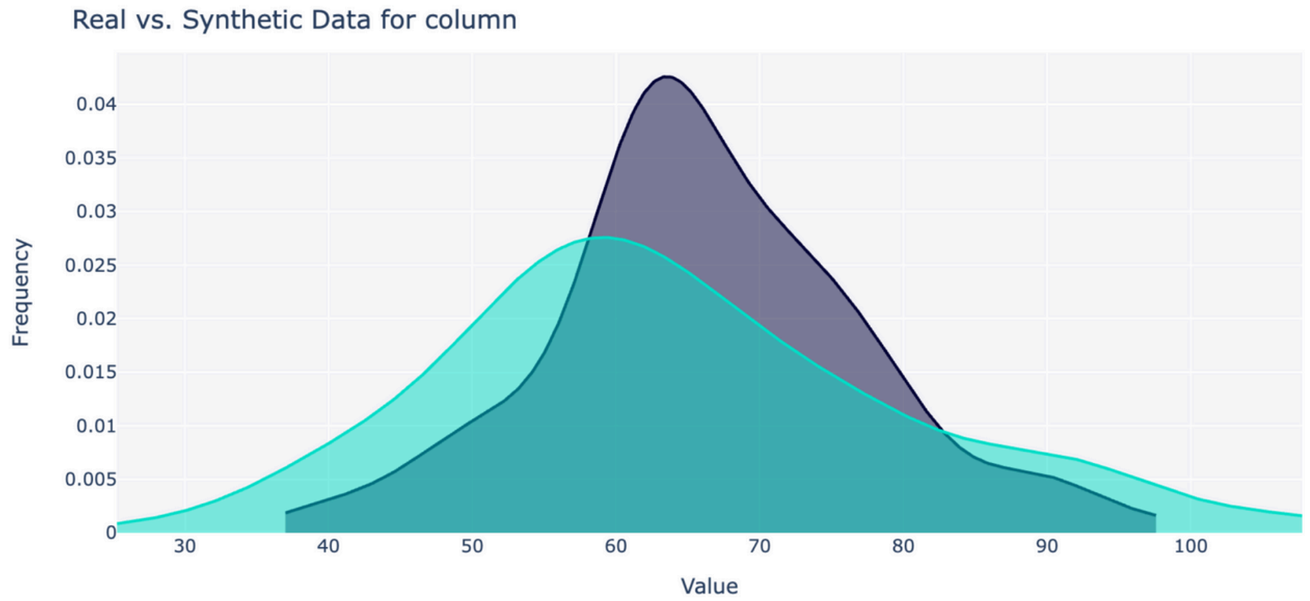
SDV

The Synthetic Data Vault

datacebo

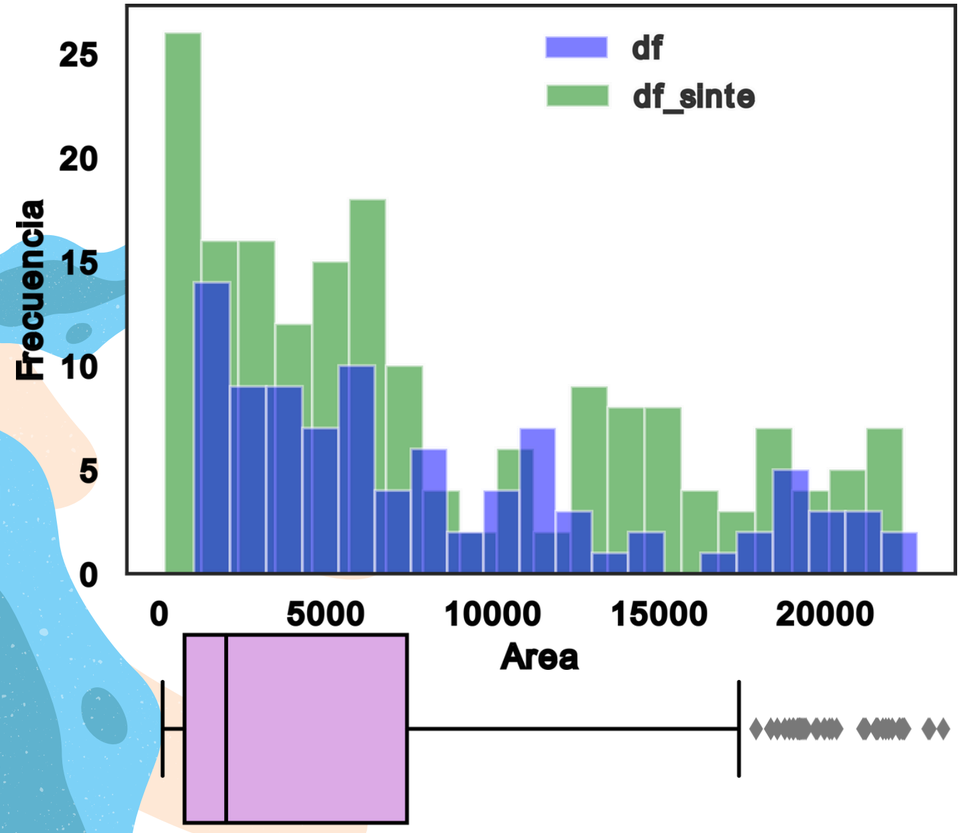
Copulas

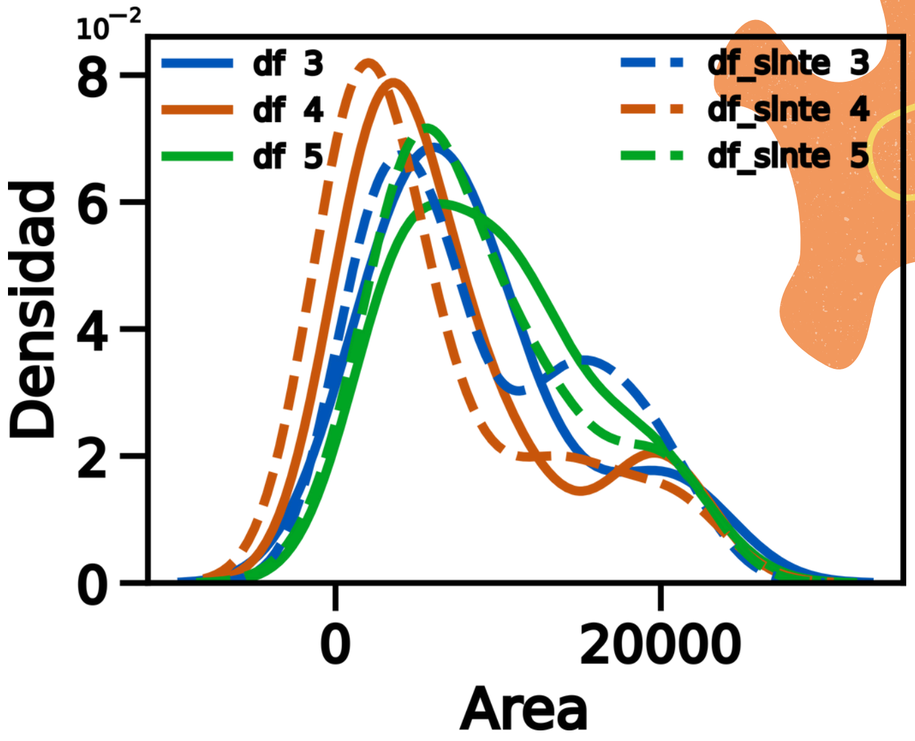
datacebo



Pocos datos!!

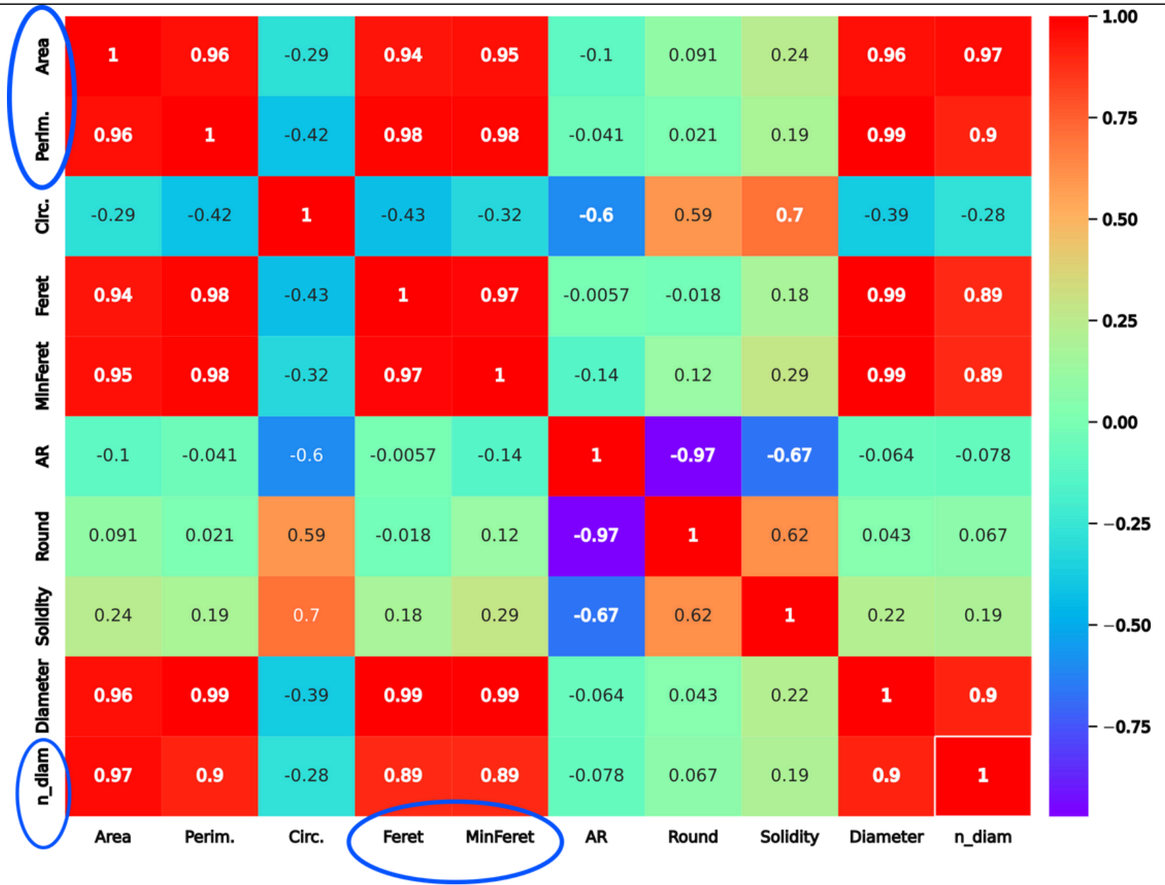
	labels	Area	Perim.	Circ.	Feret	MinFeret	AR	Round	Solidity	Esferoide	dia	Diameter	n_diam
0	0001u	2108.86	158.552	0.853148	52.0548	39.89	1.25244	0.801927	0.9574	si	3	44.78	11.1182
1	0001v	5494.72	267.744	0.765975	108.689	78.1359	1.3679	0.729508	0.9424	si	5	94.0452	55.3549
2	0001w	4604.33	208.633	0.835749	79.215	78.5711	1.0075	0.992155	0.9544	si	3	74.802	24.8861
3	0001x	3293.43	184.969	0.888199	59.5354	55.6766	1.03924	0.955215	0.9792	si	3	57.8797	19.4266
4	000a0	12597.4	393.549	0.74141	115.415	98.4842	1.12671	0.888984	0.9568	si	3	108.588	412.67





df_clean y correlación de variables

labels	Esferoide	Area	Circ.	AR	Round	Solidity	Diameter	n_diam
Esferas_BT474_dia_5_well_1_100X_3_blob_8	0	1388.7	0.9	1.1	0.9	1.0	43.0	12.9
0000v	1	4576.0	0.9	1.1	0.9	1.0	103.5	96.8
Esferas_BT474_dia_4_well_1_100X_11_blob_8	0	1061.0	0.8	1.3	0.8	1.0	38.8	9.5
000tq	1	1003.1	0.9	1.1	0.9	1.0	37.8	8.3
Esferas_BT474_dia_4_well_2_100X_1_blob_11	0	3779.8	0.6	1.2	0.8	0.9	76.5	72.6
000rk	1	6178.7	0.8	1.5	0.7	1.0	81.5	165.6
Esferas_BT474_dia_4_well_2_100X_1_blob_24	0	388.6	0.9	1.1	0.9	1.0	23.1	2.0
Esferas_BT474_dia_4_well_2_100X_2_blob_1	1	2653.2	0.9	1.1	0.9	1.0	58.7	32.8
Esferas_BT474_dia_5_well_2_100X_1_blob_4	0	150.5	0.9	1.1	0.9	0.9	14.3	0.5
000cd	1	7261.1	0.8	1.0	1.0	1.0	95.5	156.5



MODELOS DE APRENDIZAJE

Modelos:

- Regresión Logística
- Árbol de Decisión
- Random Forest
- XGboost

Datos:

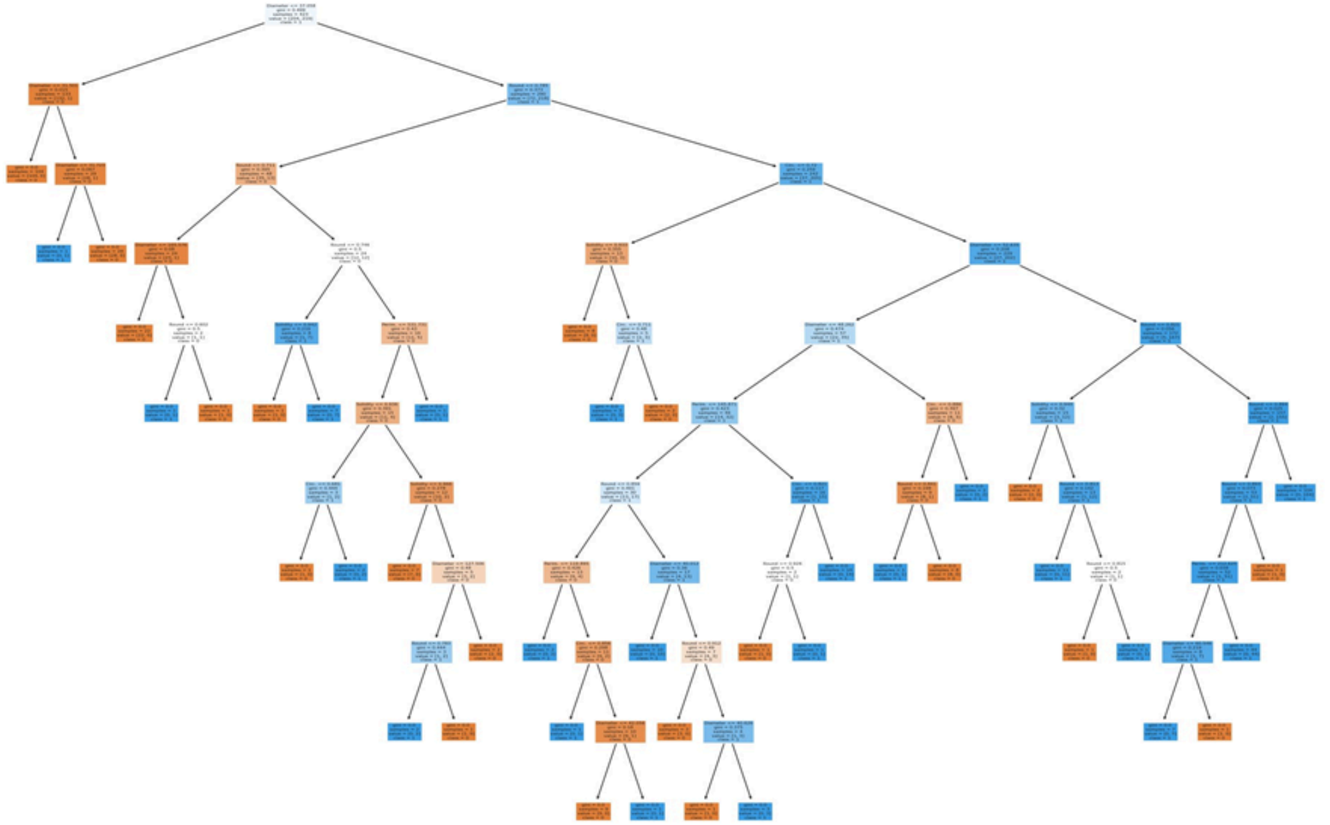
- 80% entrenamiento
- 20% testeo
- 3-4-5 días

Variables:

- Diámetro
- Round
- Perim.
- Circ.
- Solidity

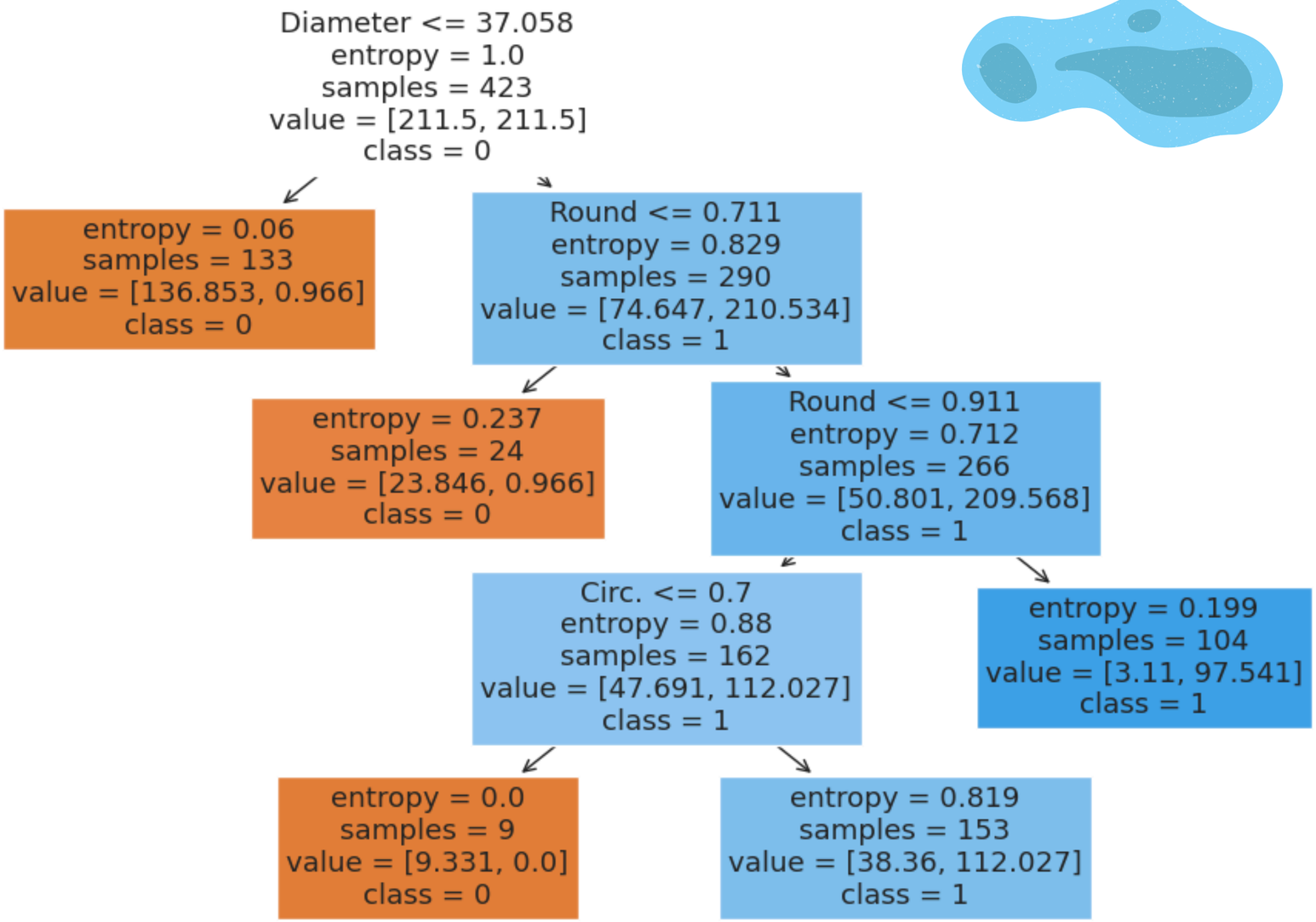
Árbol de Decisión Base

Se observó algunas variables como: diámetro, redondez y circularidad, tiene mayor probabilidad de ser clasificadas como tumoresferas.



Árbol de Decisión Ajustado

Se observó que las células con un diámetro mayor a 37.058, una redondez superior a 0.711 y una circunferencia mayor a 0.7 tienen una mayor probabilidad de ser clasificadas como Tumoresferas. Estos datos son valiosos para identificar características específicas que podrían indicar la presencia de Tumoresferas en futuras observaciones.

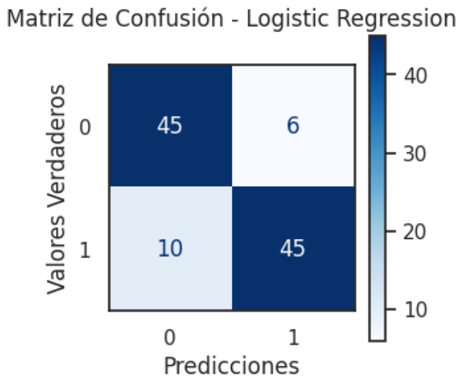


Regresión Logística Base

Informe de Clasificación:

	precision	recall	f1-score	support
0	0.82	0.88	0.85	51
1	0.88	0.82	0.85	55
accuracy			0.85	106
macro avg	0.85	0.85	0.85	106
weighted avg	0.85	0.85	0.85	106

Accuracy Score (Puntuación de Exactitud):0.85
Precision Score (Puntuación de Precisión):0.88
F1 Score:0.85

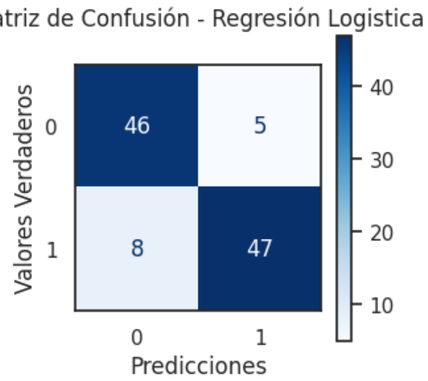


Regresión Logística Ajustada

Informe de Clasificación:

	precision	recall	f1-score	support
0	0.85	0.90	0.88	51
1	0.90	0.85	0.88	55
accuracy			0.88	106
macro avg	0.88	0.88	0.88	106
weighted avg	0.88	0.88	0.88	106

Accuracy Score (Puntuación de Exactitud): 0.88
Precision Score (Puntuación de Precisión): 0.9
F1 Score: 0.88



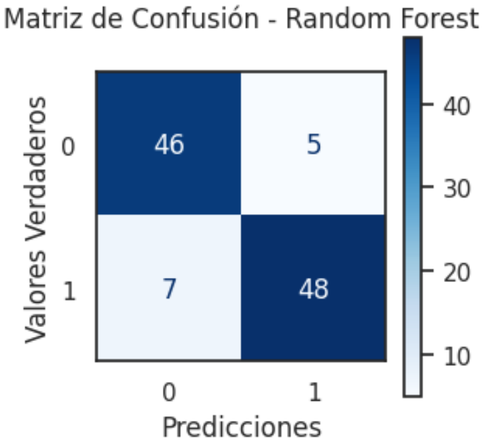
RandomForest Base

Random Forest Sin Ajustar:

Informe de Clasificación:

	precision	recall	f1-score	support
0	0.88	0.90	0.89	51
1	0.91	0.89	0.90	55
accuracy			0.90	106
macro avg	0.90	0.90	0.90	106
weighted avg	0.90	0.90	0.90	106

Accuracy Score (Puntuación de Exactitud):0.9
Precision Score (Puntuación de Precisión):0.91
F1 Score:0.9



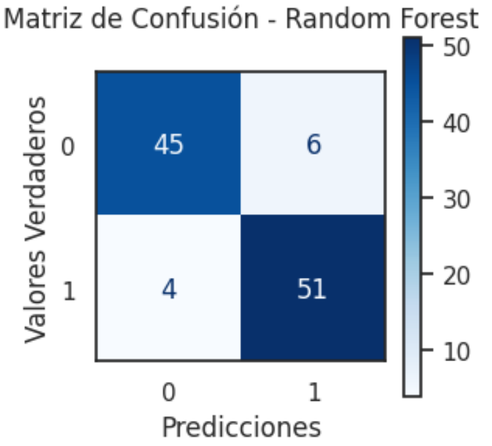
RandomForest Ajustada

Random Forest Ajustado:

Informe de Clasificación:

	precision	recall	f1-score	support
0	0.92	0.88	0.90	51
1	0.89	0.93	0.91	55
accuracy			0.91	106
macro avg	0.91	0.90	0.91	106
weighted avg	0.91	0.91	0.91	106

Accuracy Score (Puntuación de Exactitud): 0.91
Precision Score (Puntuación de Precisión): 0.89
F1 Score: 0.91



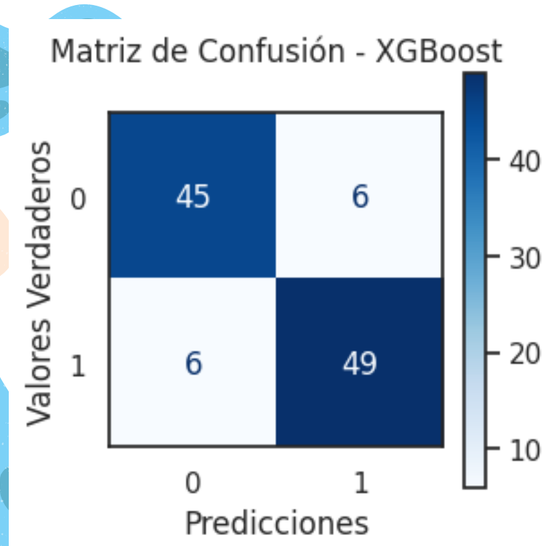
XGboost Base

XGboost Sin Ajustar:

Informe de Clasificación:

	precision	recall	f1-score	support
0	0.88	0.88	0.88	51
1	0.89	0.89	0.89	55
accuracy			0.89	106
macro avg	0.89	0.89	0.89	106
weighted avg	0.89	0.89	0.89	106

Accuracy Score (Puntuación de Exactitud):0.89
Precision Score (Puntuación de Precisión):0.89
F1 Score:0.89

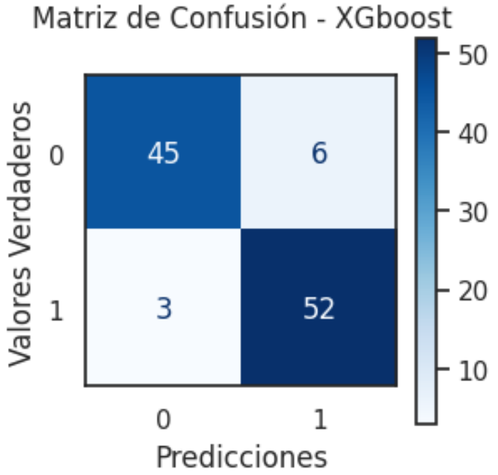


XGboost Ajustada

Informe de Clasificación:

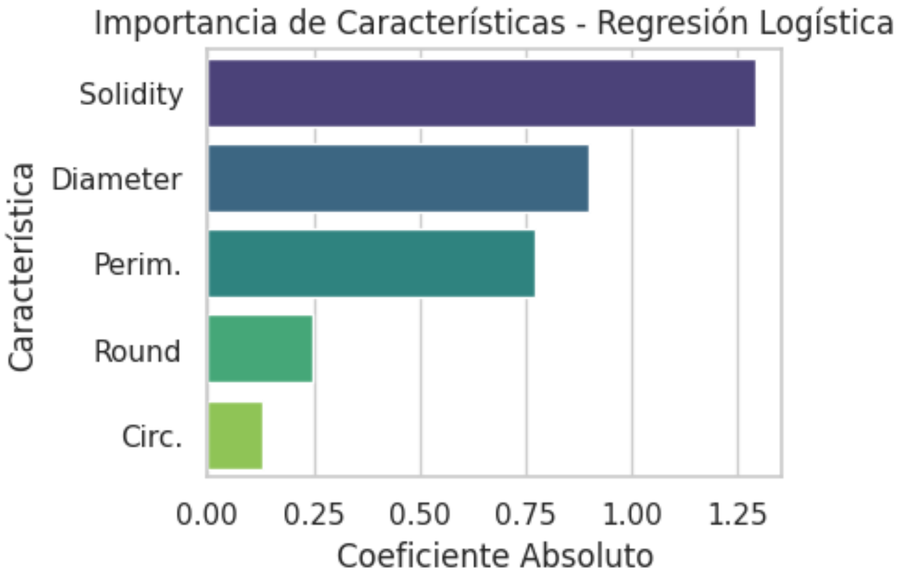
	precision	recall	f1-score	support
0	0.94	0.88	0.91	51
1	0.90	0.95	0.92	55
accuracy			0.92	106
macro avg	0.92	0.91	0.91	106
weighted avg	0.92	0.92	0.91	106

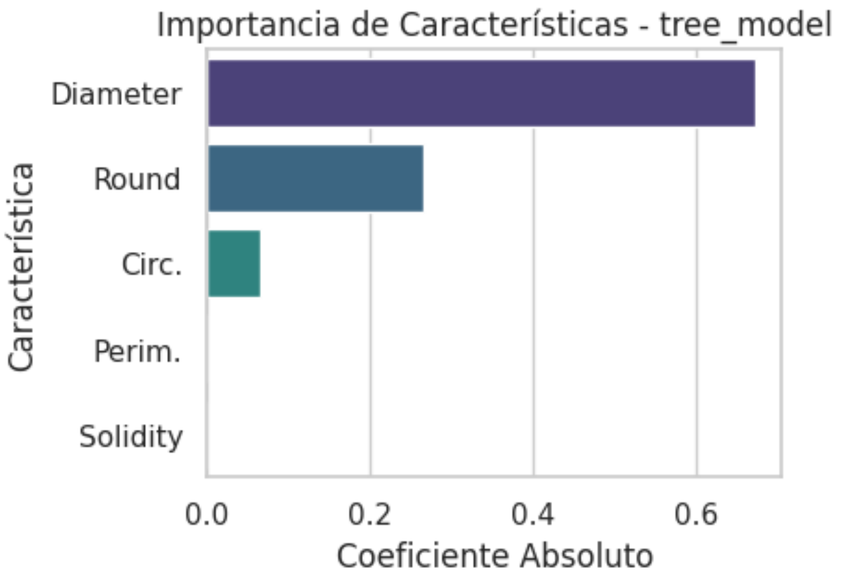
Accuracy Score (Puntuación de Exactitud): 0.92
Precision Score (Puntuación de Precisión): 0.9
F1 Score: 0.92

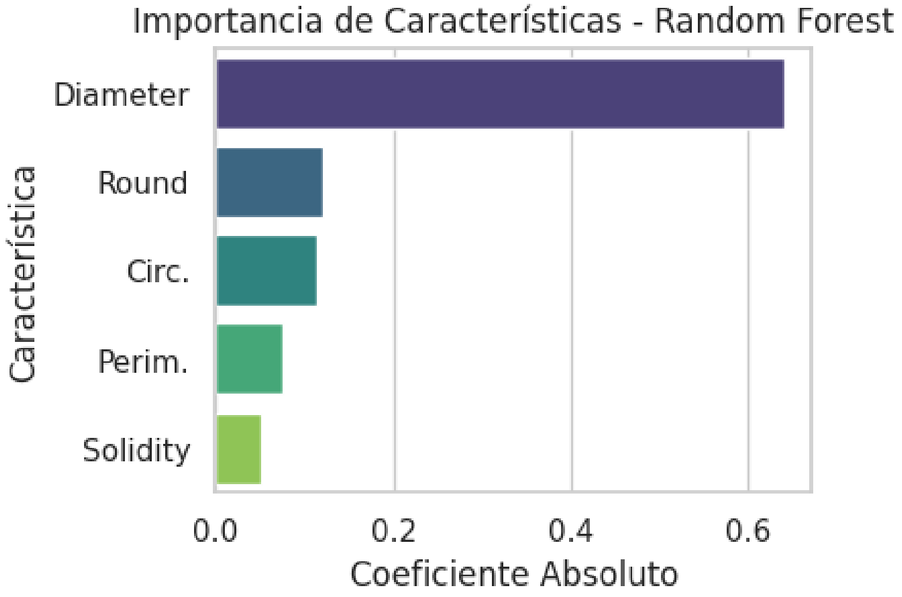


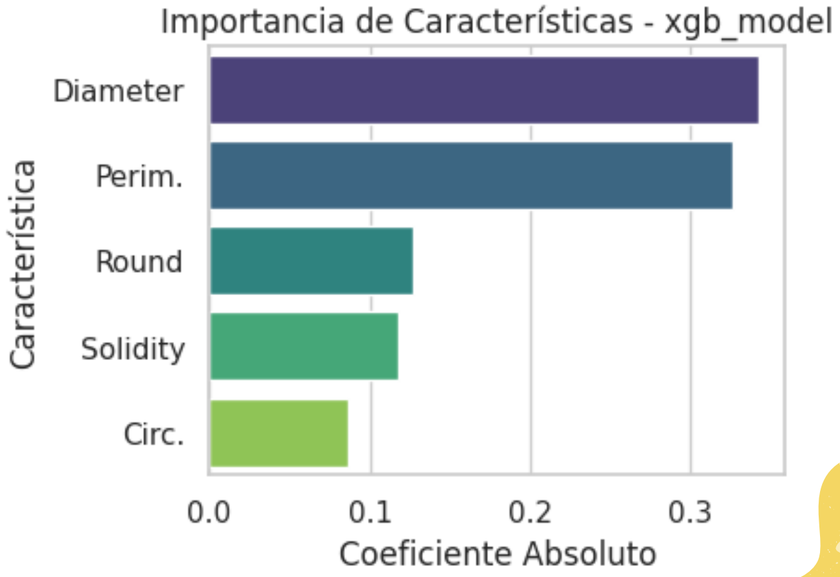
Características importantes para los modelos

En términos generales, para los cuatros modelos, se puede destacar la relevancia de la variable a diámetro al momento de ajuste del modelo y denota la posible importancia que esta variable podría tener en la clasificación de las tumoresferas y no tumoresferas.









CONCLUSIONES Y OBSERVACIONES

Modelos Base vs. Modelos Ajustados

En general, los modelos ajustados con hiperparámetros superan en términos de Accuracy y F1-score a los modelos de base, lo que indica que el ajuste puede mejorar significativamente el rendimiento de los modelos.

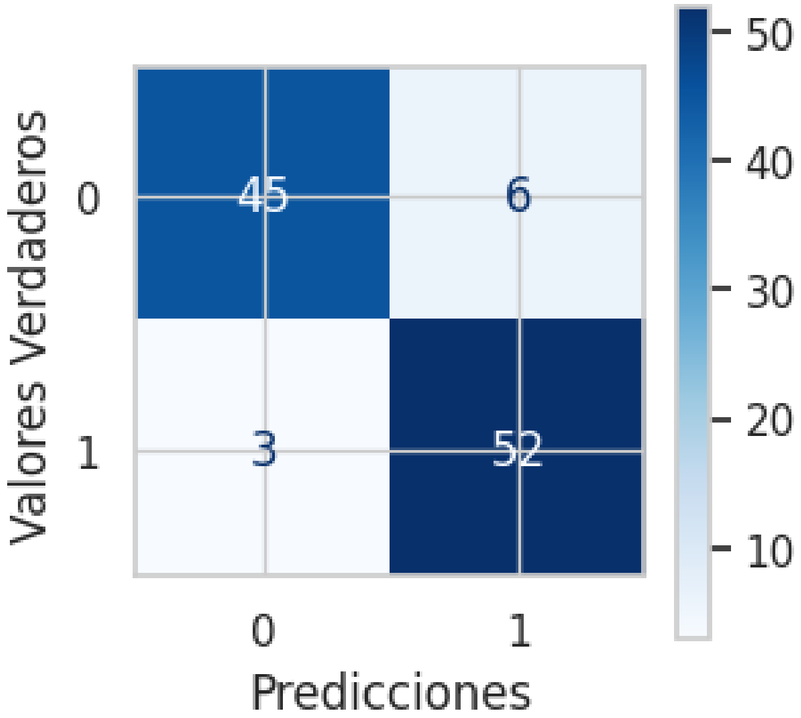
Modelos Base

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.849057	0.882353	0.818182	0.849057
Decision Tree	0.839623	0.895833	0.781818	0.834951
Random Forest	0.896226	0.907407	0.890909	0.899083
XGBoost	0.886792	0.890909	0.890909	0.890909

Modelos Ajustados

Model	Accuracy	Precision	Recall	F1 Score
LogisticRegression	0.877358	0.903846	0.854545	0.878505
DecisionTreeClassifier	0.915094	0.896552	0.945455	0.920354
Random Forest	0.905660	0.894737	0.927273	0.910714
XGBoost	0.915094	0.896552	0.945455	0.920354

XGboost – D.Tree




```
Informe de Clasificación:

              precision    recall  f1-score   support

           0       0.94      0.88      0.91        51
           1       0.90      0.95      0.92        55

    accuracy                           0.92       106
   macro avg       0.92      0.91      0.91       106
weighted avg       0.92      0.92      0.91       106


Accuracy Score (Puntuación de Exactitud): 0.92
Precision Score (Puntuación de Precisión): 0.9
F1 Score: 0.92
```

- **Accuracy (Exactitud)**

El 92% de las predicciones son correctas.

- **Precision (Precisión)**

El 90% de las predicciones de tumoresferas fueron correctas

- **Recall (Recuperación o Sensibilidad)**

El modelo detecto correctamente el 95% de las tumoresferas reales.

- **F1-Score**

Un valor de 92% indica un buen equilibrio entre la precisión y la capacidad del modelo para detectar tumoresferas.

Experiencia

Gracias!!