

# The iris data set

Martín Quijano – Martina Coletto

## Variables

- Sepal Length
- Sepal Width
- Petal Length
- Petal Width

## Modelos

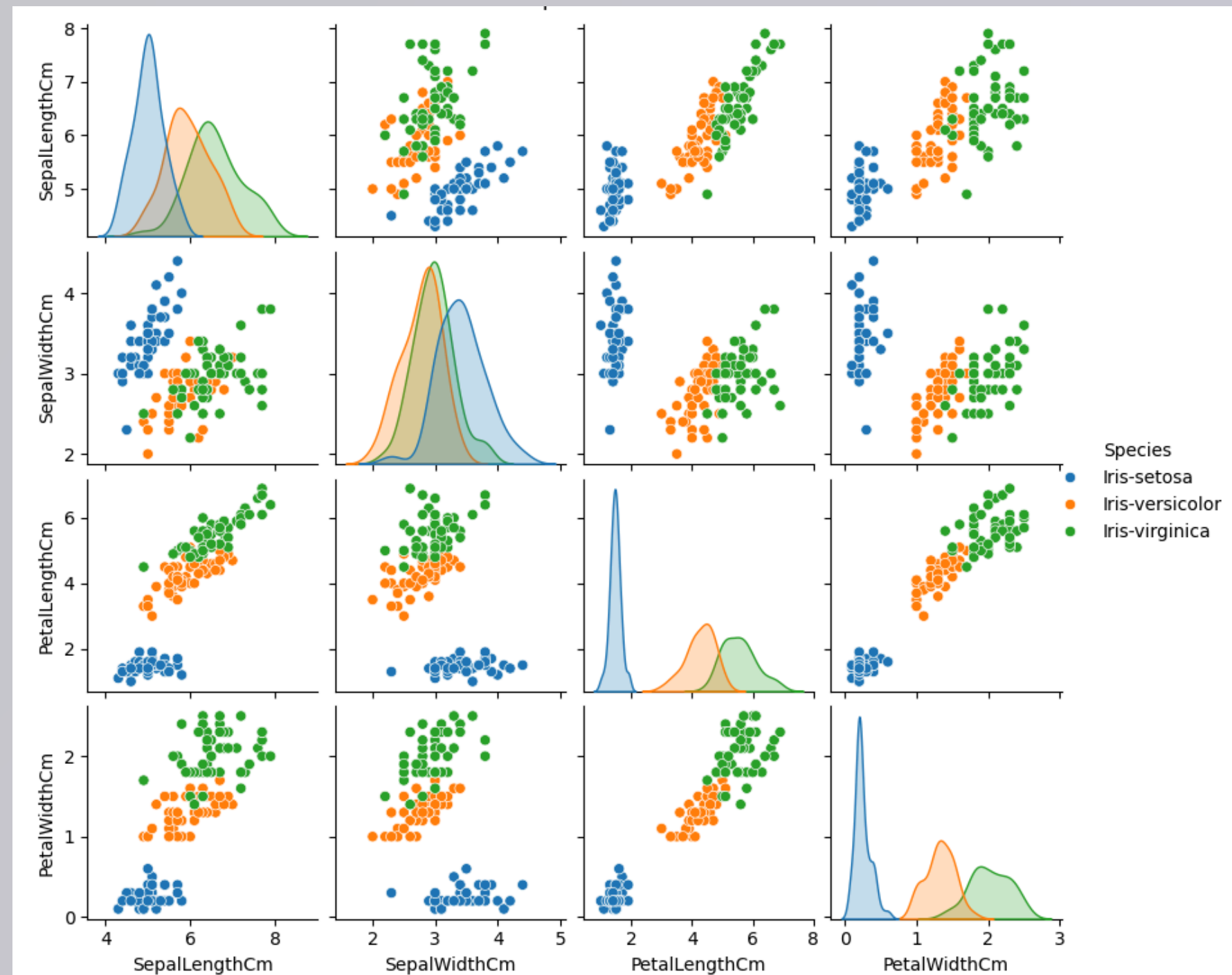
K-Nearest Neighbors (KNN) y Regresión.

## Objetivo

Clasificar las flores del dataset Iris utilizando aprendizaje supervisado.

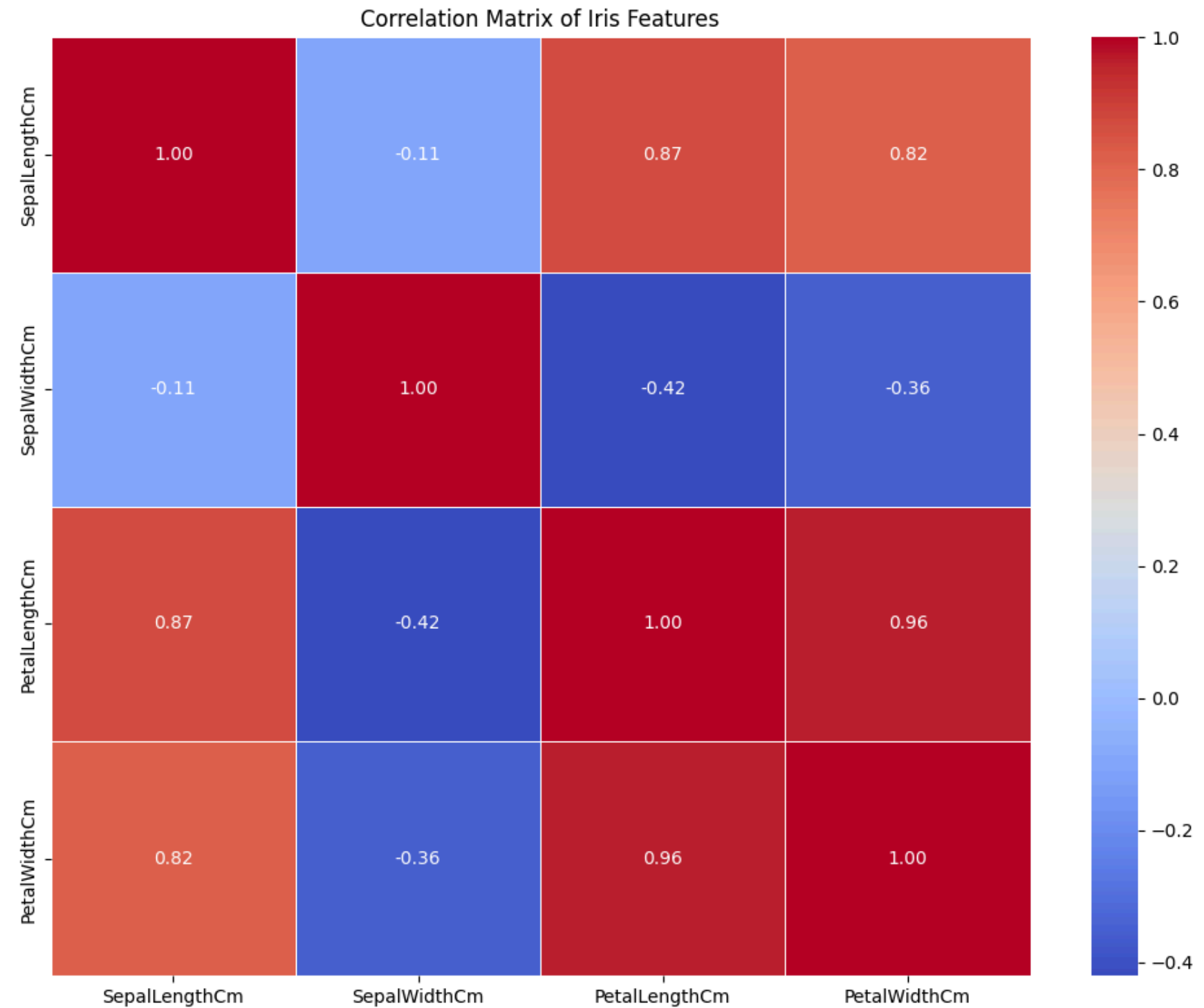
# Pairplot

Para observar la distribución y posibles relaciones entre las variables.



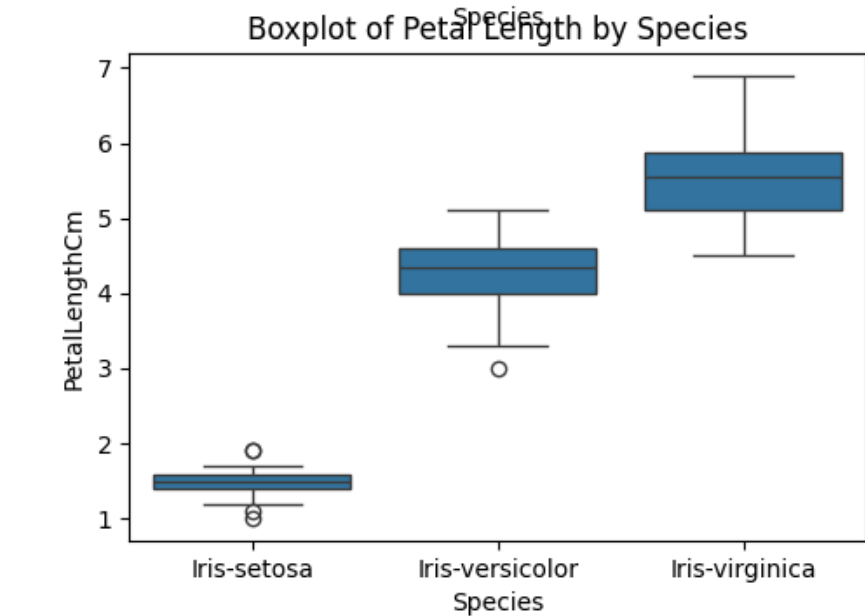
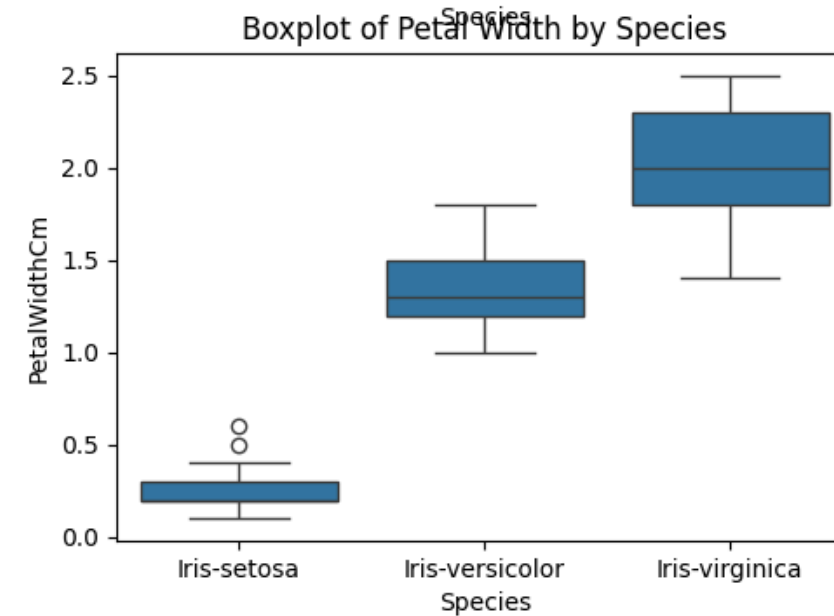
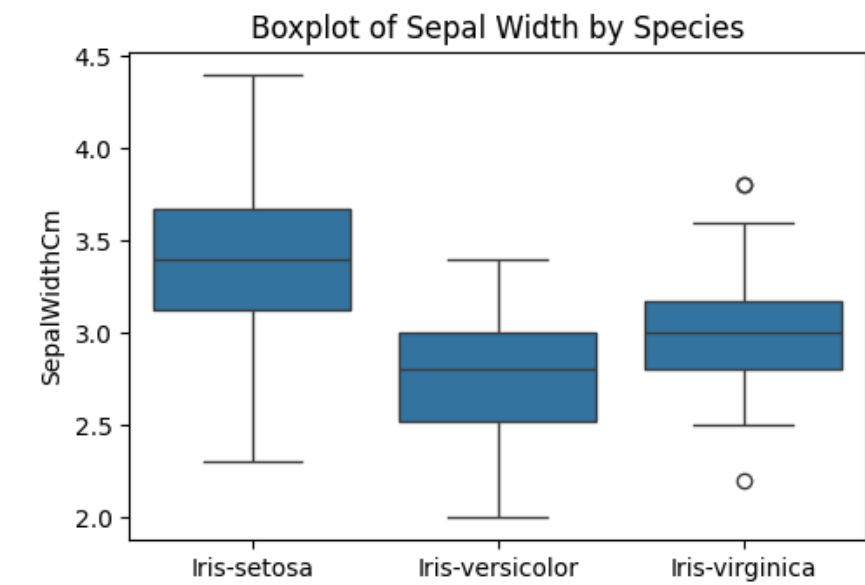
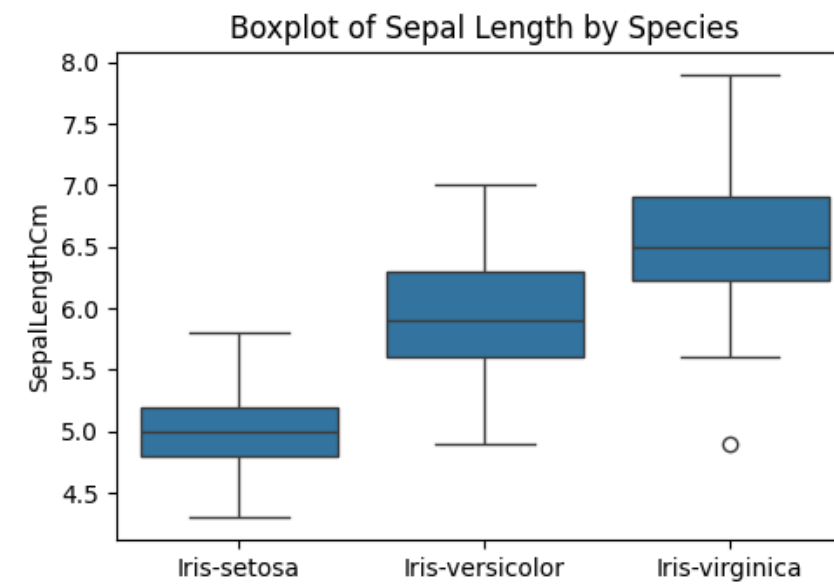
# Matriz de Correlación

Para identificar dependencias entre las variables.

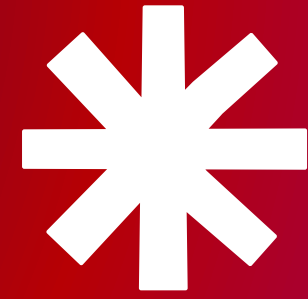


# Boxplot

Para detectar outliers y diferencias en la distribución de las clases.



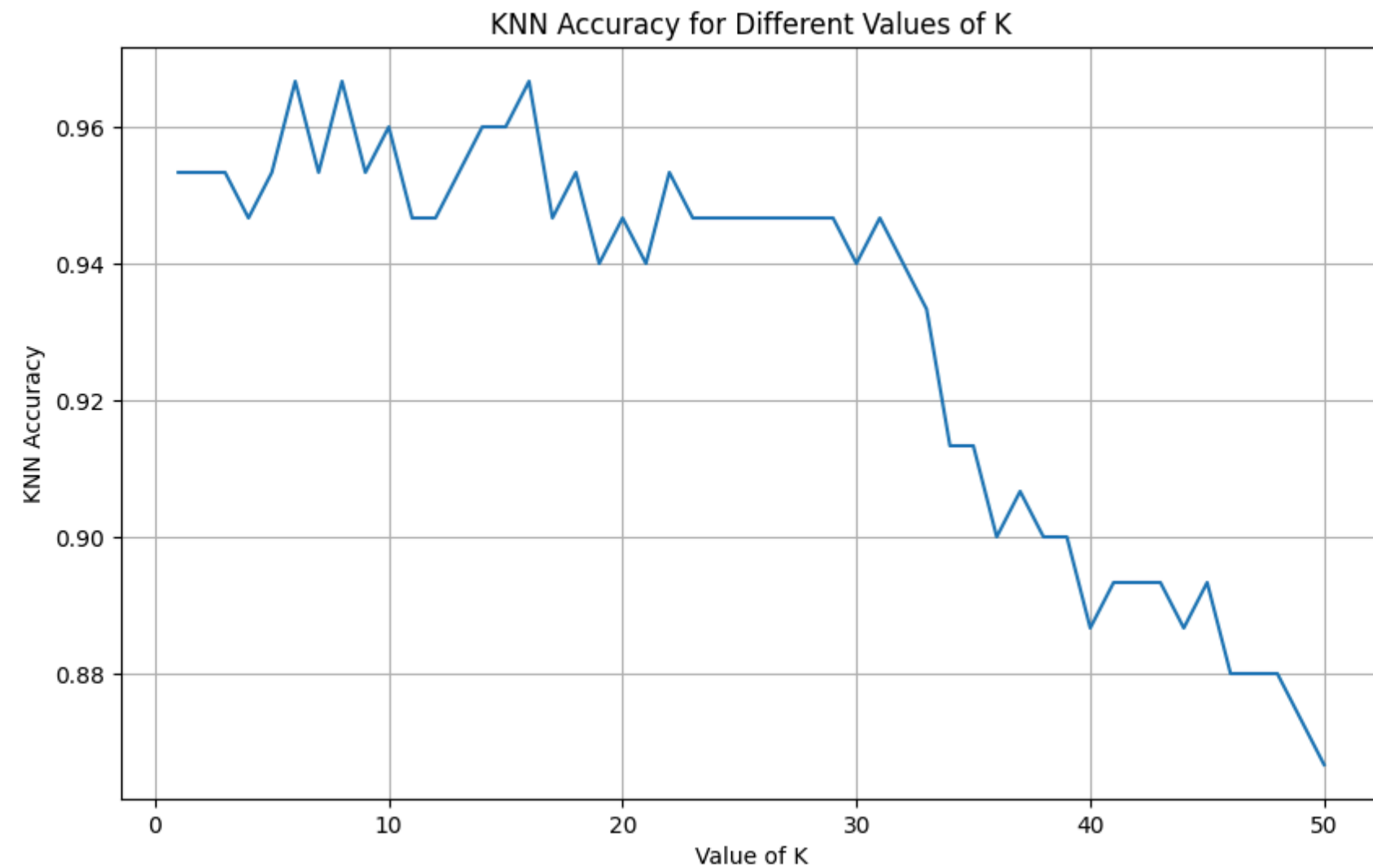




# Análisis con KNN

Elección del mejor valor de K  
mediante cross validation.

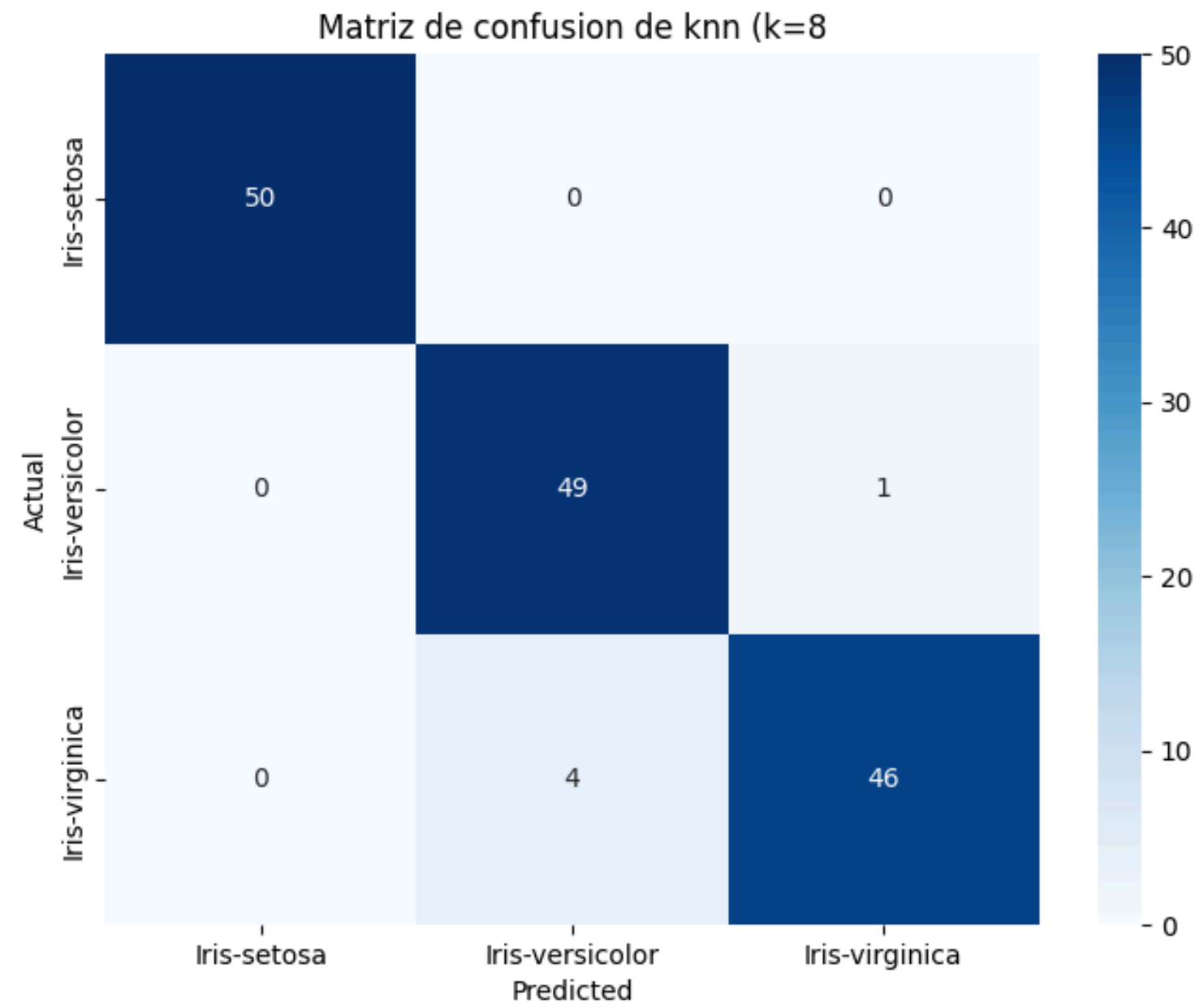
**Accuracy: 0.9667**

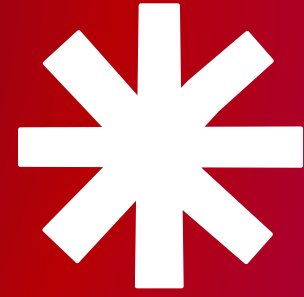




# Análisis con KNN

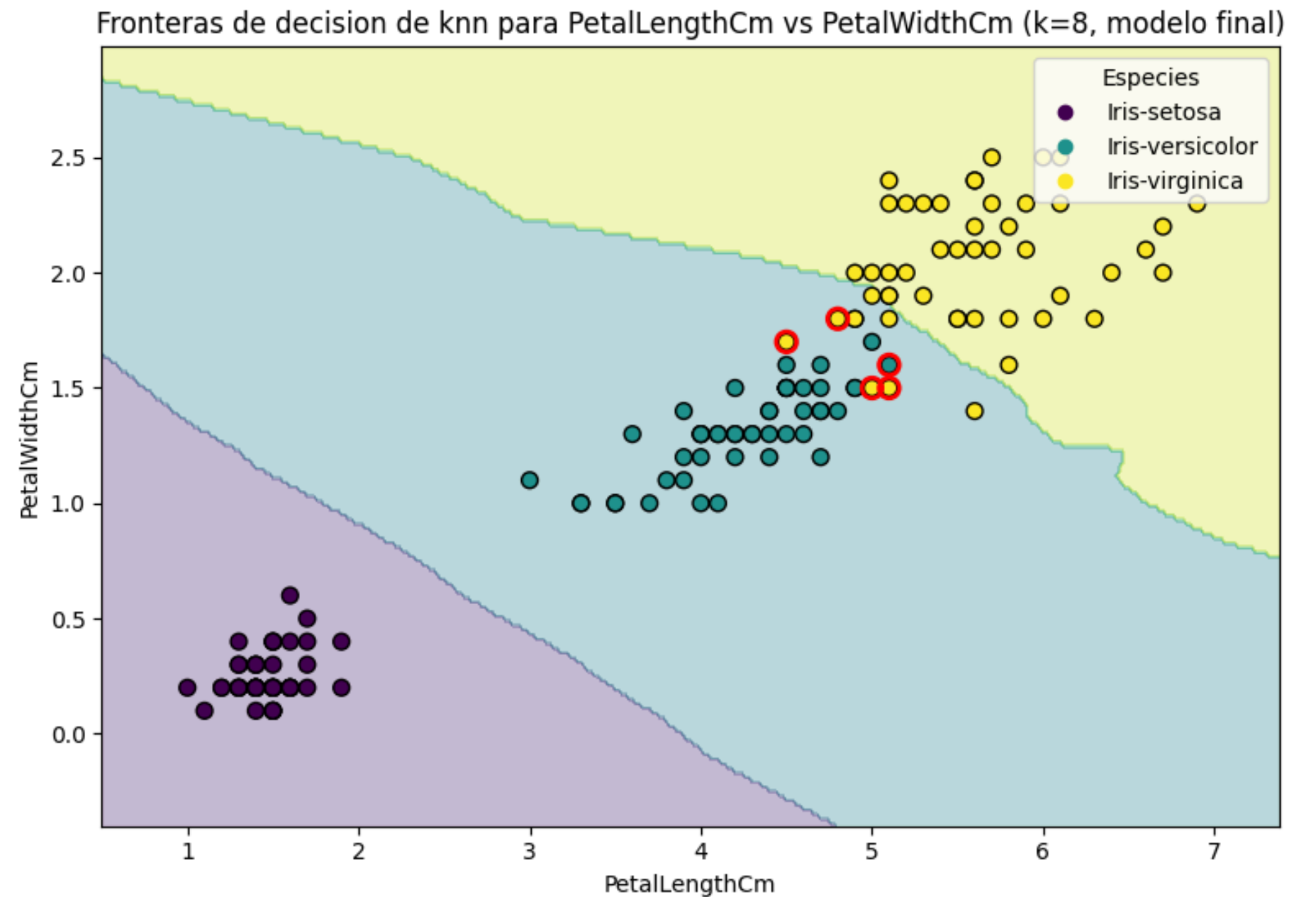
Matriz de confusión





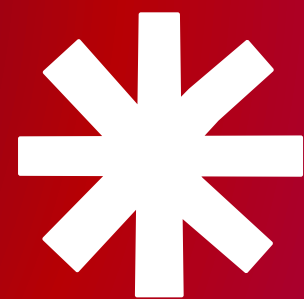
# Análisis con KNN

Fronteras de decisión



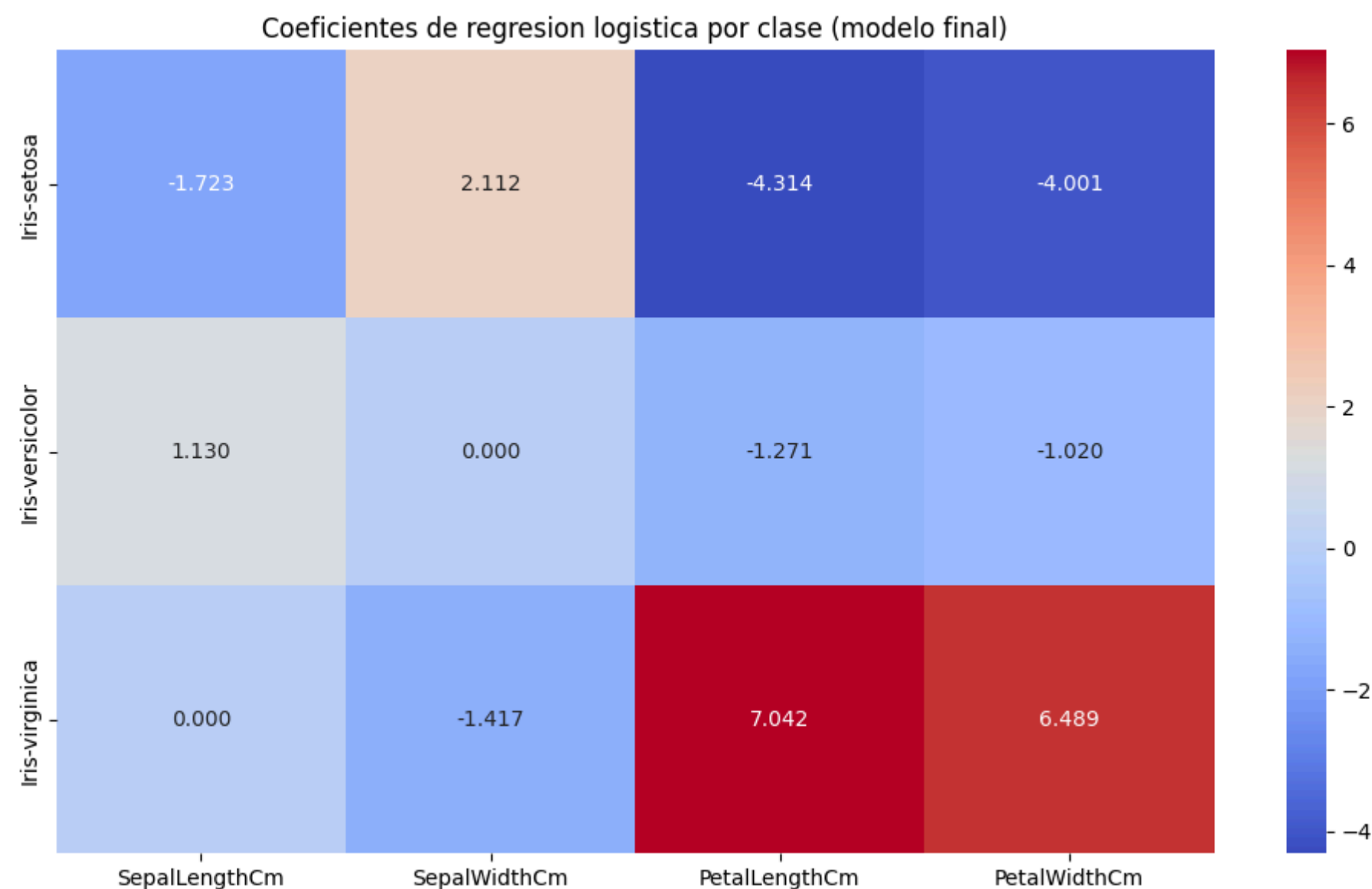
Nota: Los colores de los puntos muestran las etiquetas verdaderas, los colores de fondo muestran las predicciones del modelo.  
Los círculos rojos indican los puntos mal clasificados.





# Análisis de Regresión Logística

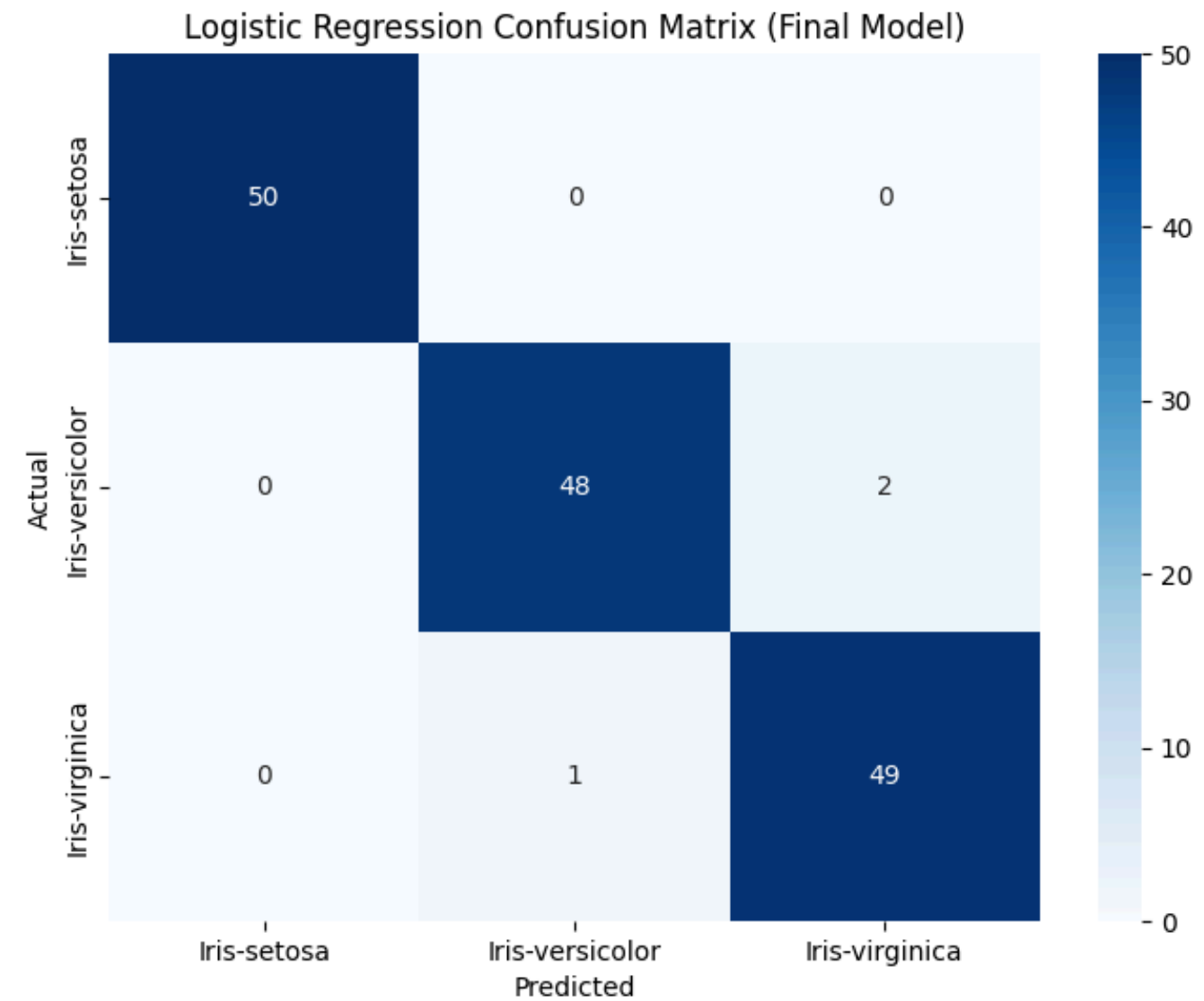
Evaluación de betas.  
**Accuracy: 0.98**

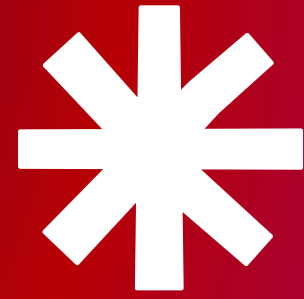




# Análisis de Regresión Logística

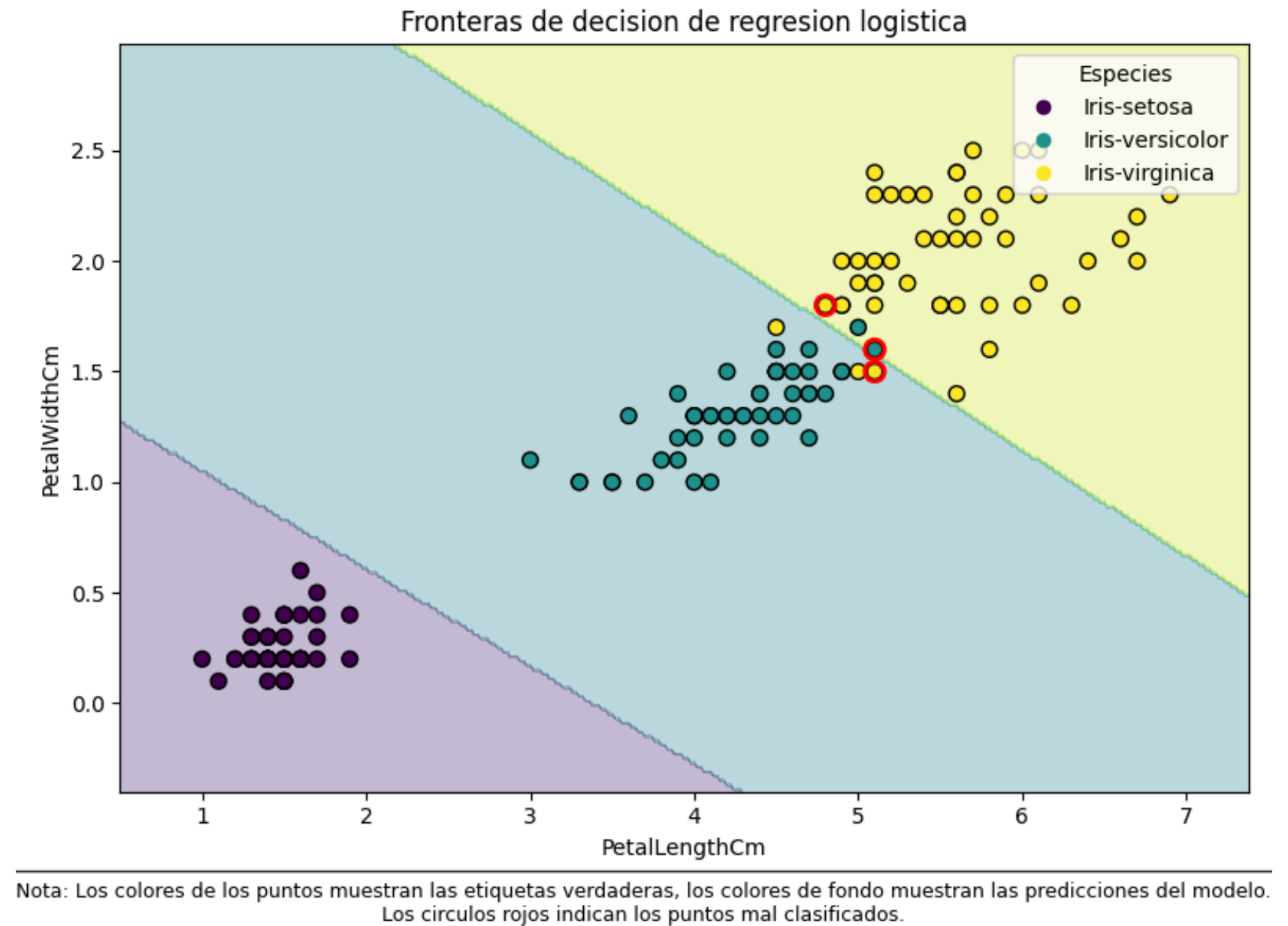
Matriz de confusión





# Análisis de Regresión Logística

Fronteras de decision



# Accuracy de los modelos con las transformaciones

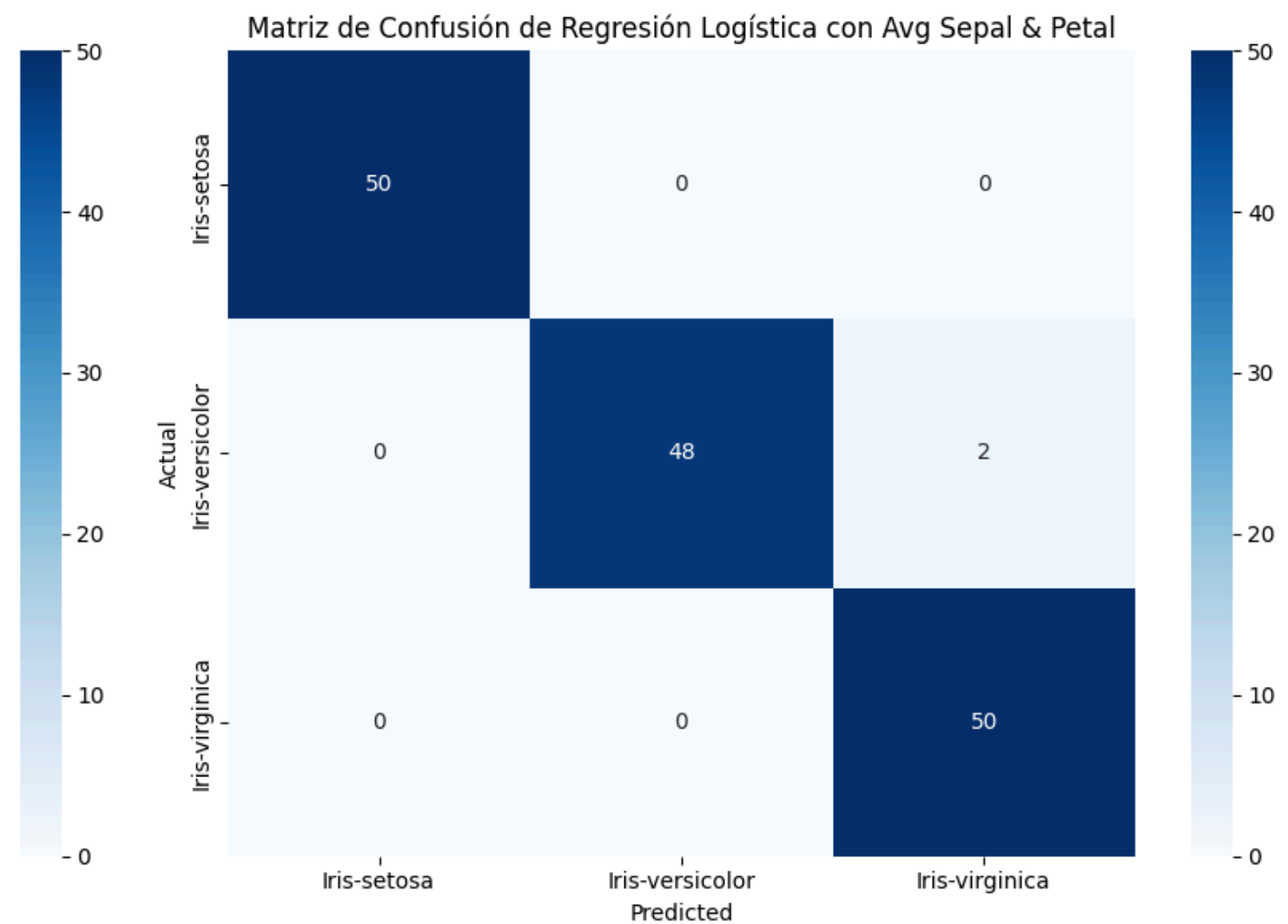
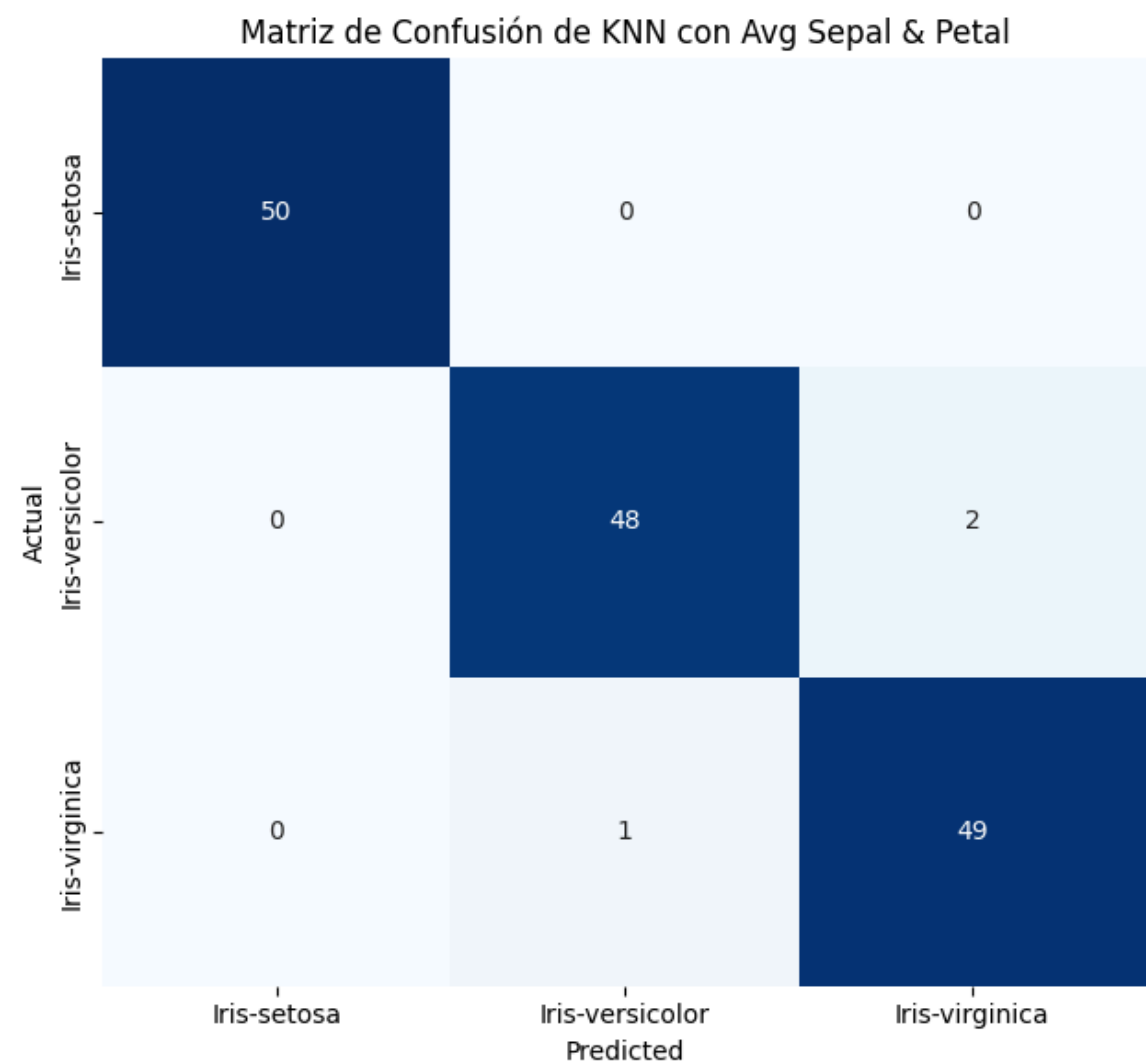
accuracy de KNN por Transformación

Transformación	accuracy	
MinMax Scaling	0.9733	★
Avg Sepal & Petal	0.9733	★
Original	0.9667	
Ratios	0.9667	
Logarithmic	0.9600	
Avg Height & Width	0.9267	

accuracy de Regresión Logística por Transformación

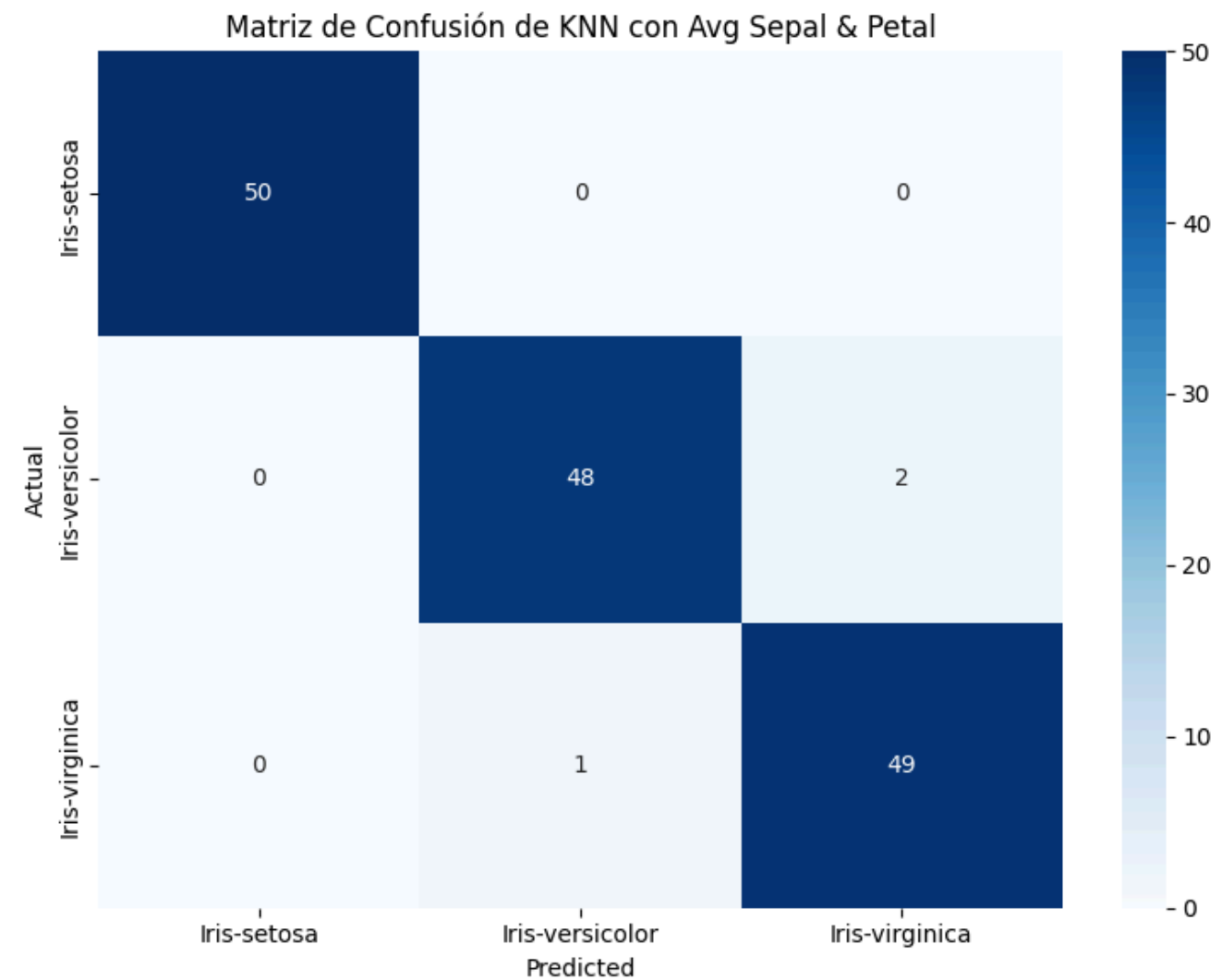
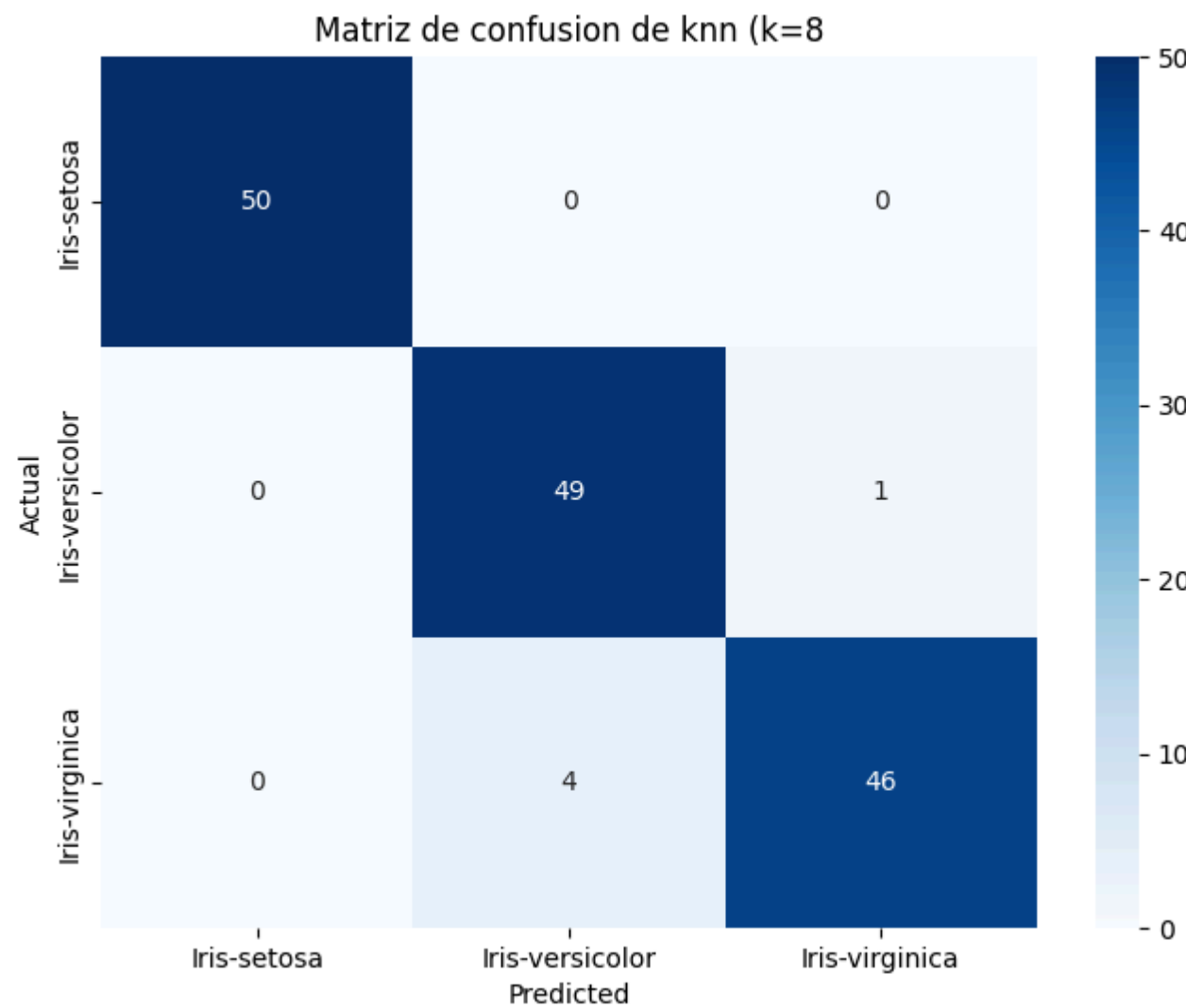
Transformación	accuracy	
Avg Sepal & Petal	0.9867	★
Original	0.9800	
MinMax Scaling	0.9800	
Logarithmic	0.9800	
Ratios	0.9600	
Avg Height & Width	0.9000	

## Matriz de confusión de los modelos con mejor transformación

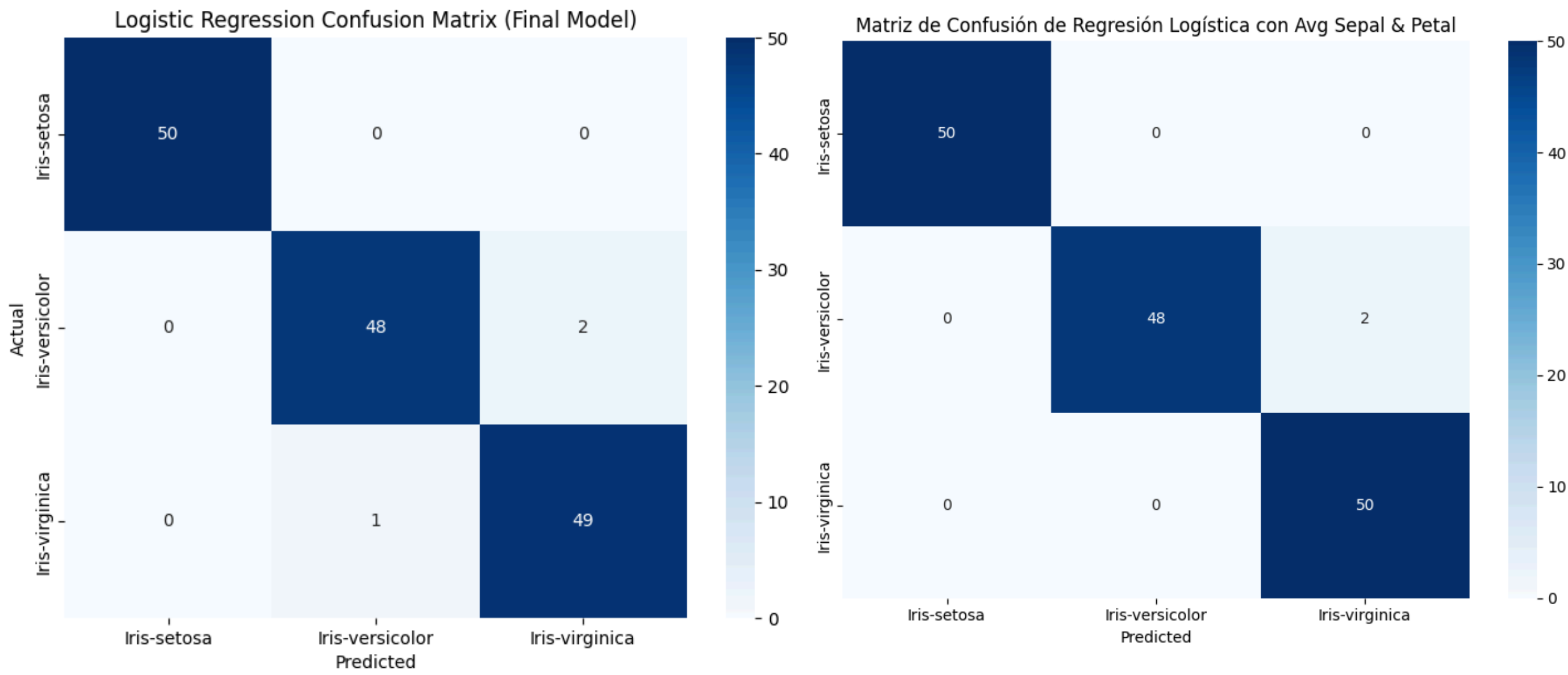




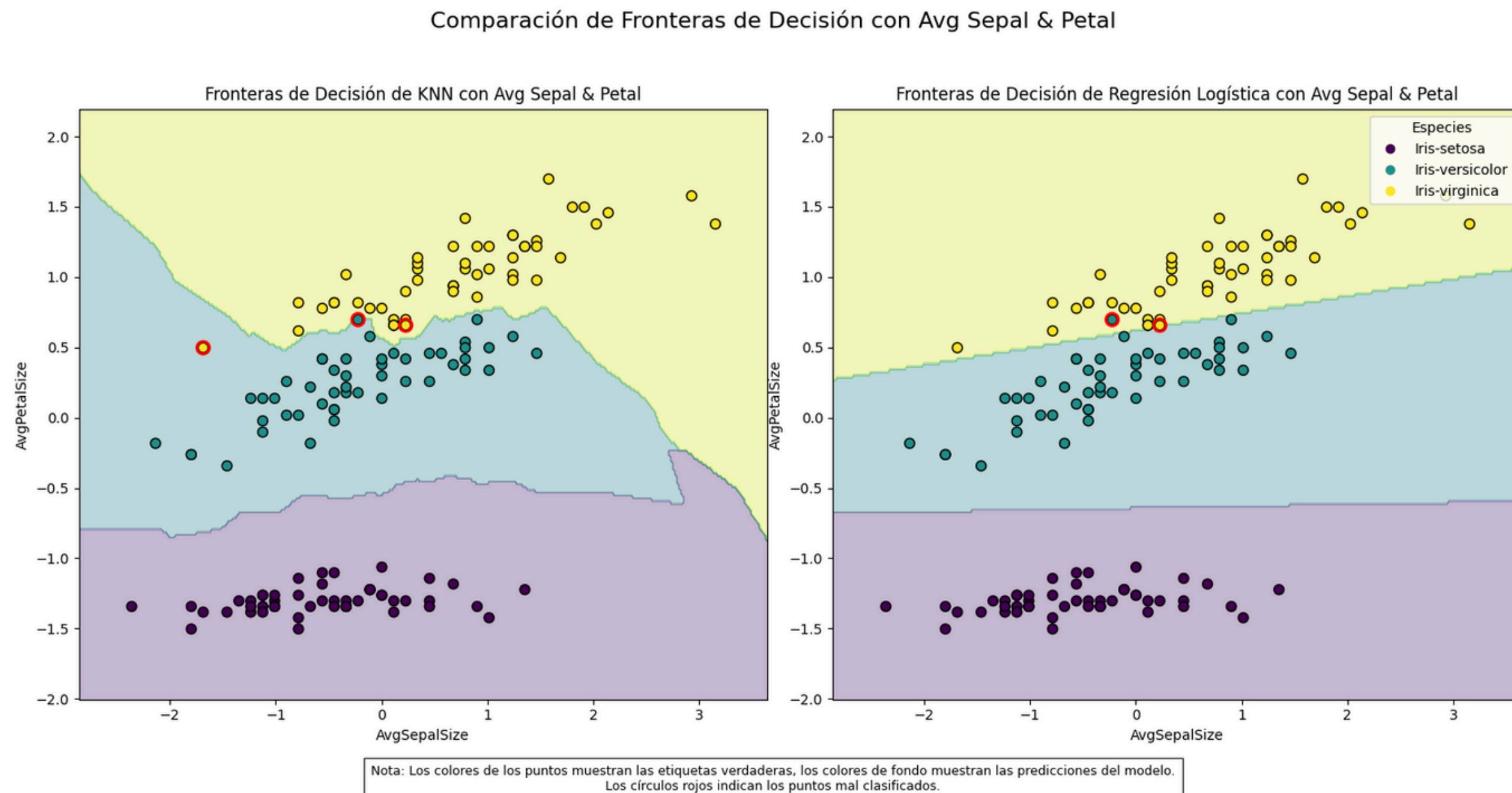
## KNN: normal vs. transformación



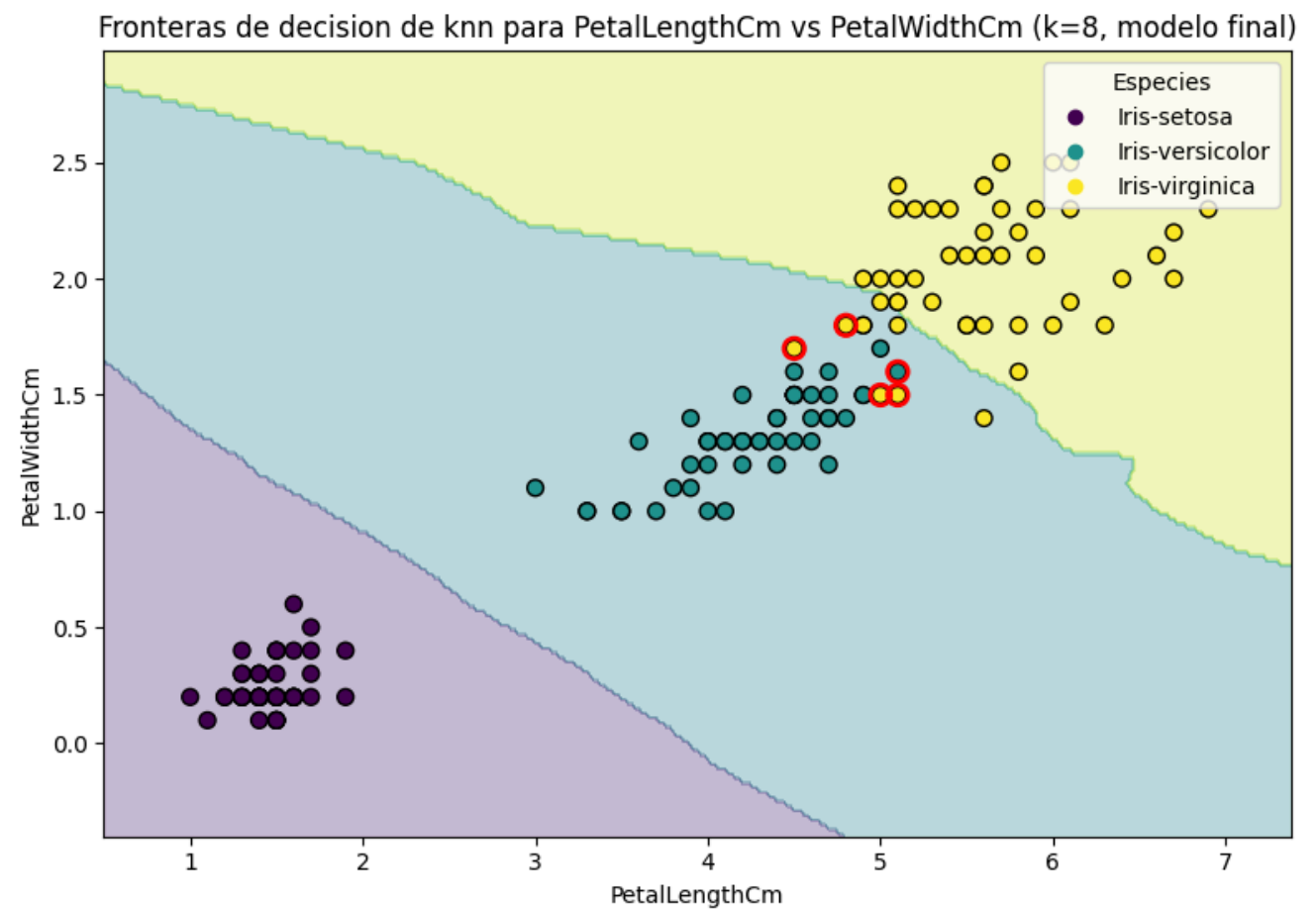
# Regresión Logística: normal vs. transformación



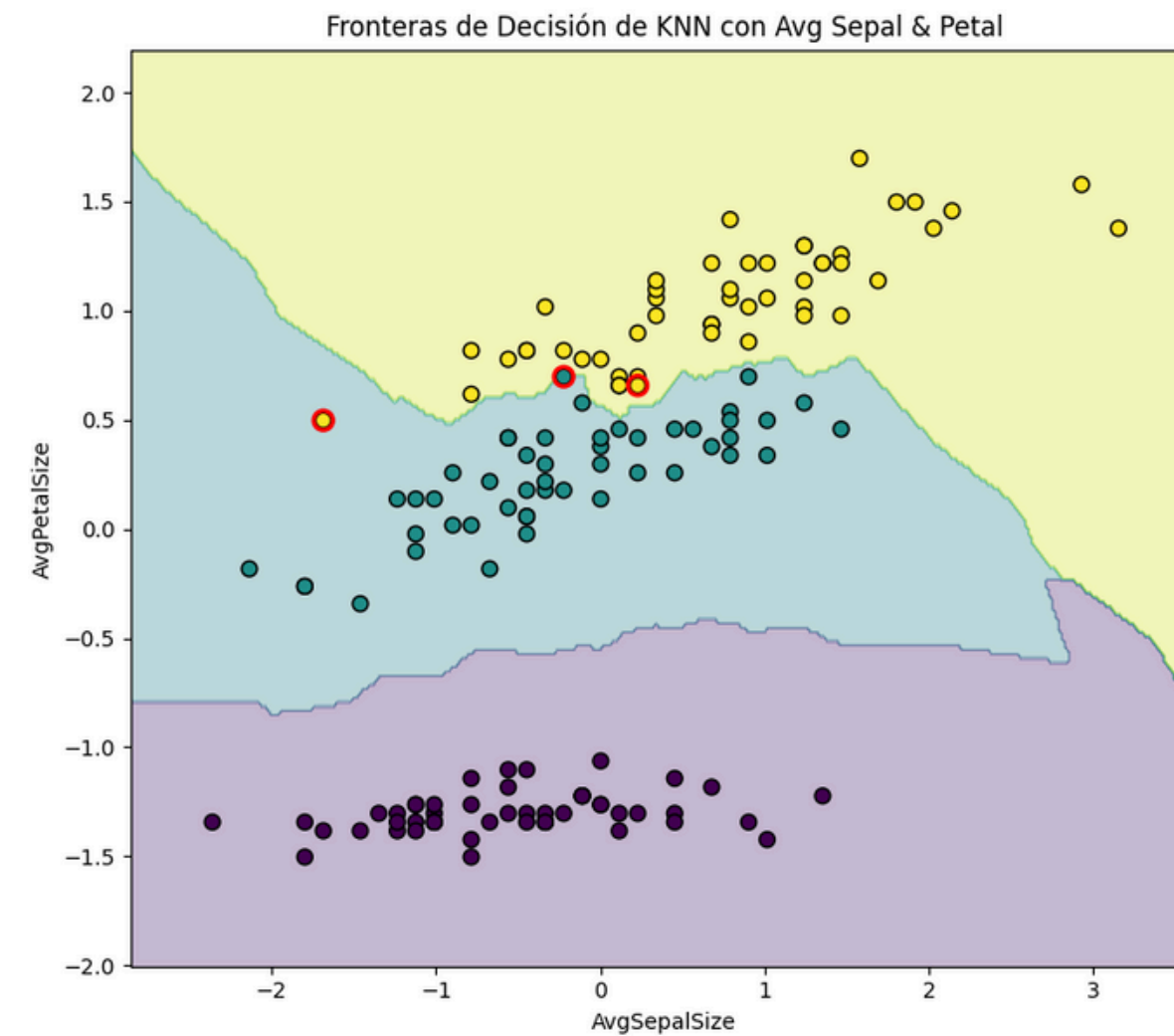
## Fronteras de decisión de los modelos con mejor transformación



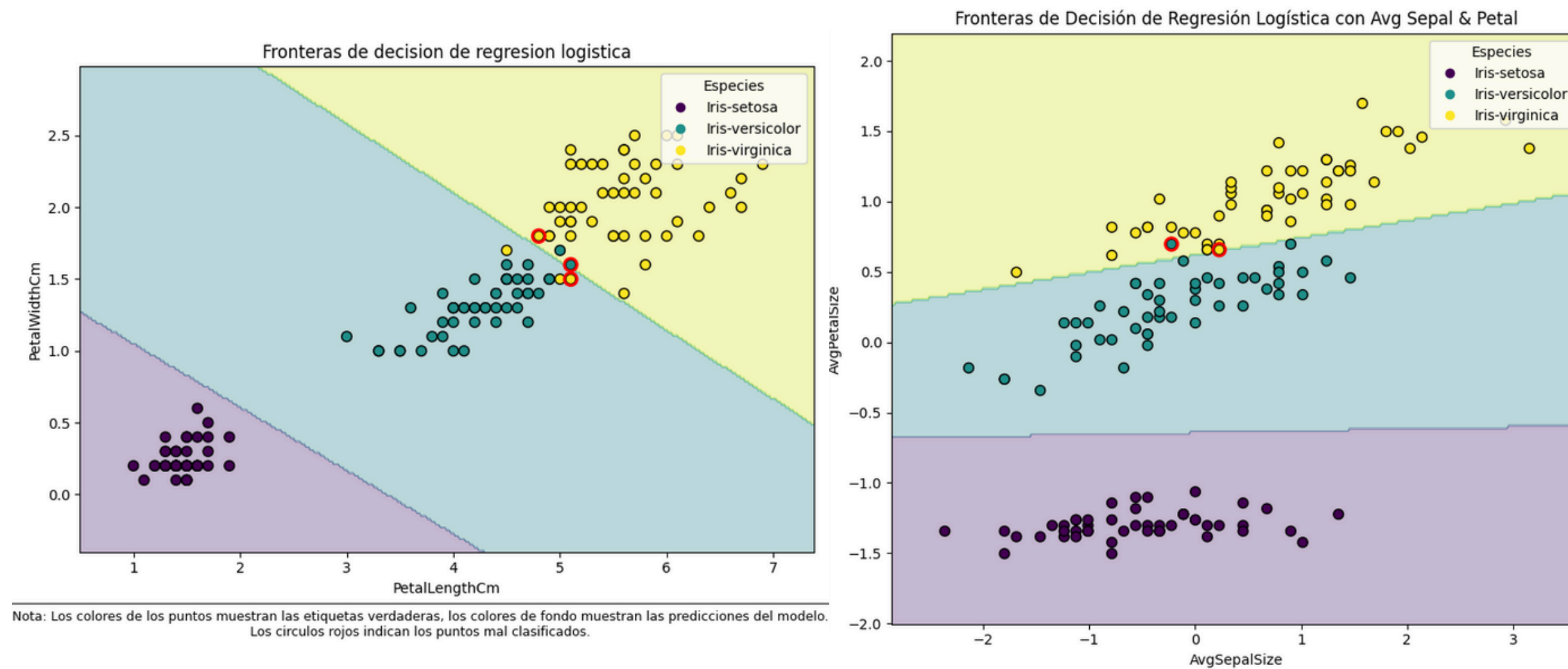
## KNN: normal vs. transformación



Nota: Los colores de los puntos muestran las etiquetas verdaderas, los colores de fondo muestran las predicciones del modelo.  
Los círculos rojos indican los puntos mal clasificados.



## Regresión Logística: normal vs. transformación





## ¿Las observaciones son agrupables por especie?

Si, como vemos en la matriz de confusión y en las fronteras de decisión las observaciones son agrupables con un margen de error muy chico.

# Gracias

Martín Quijano – Martina Coletto