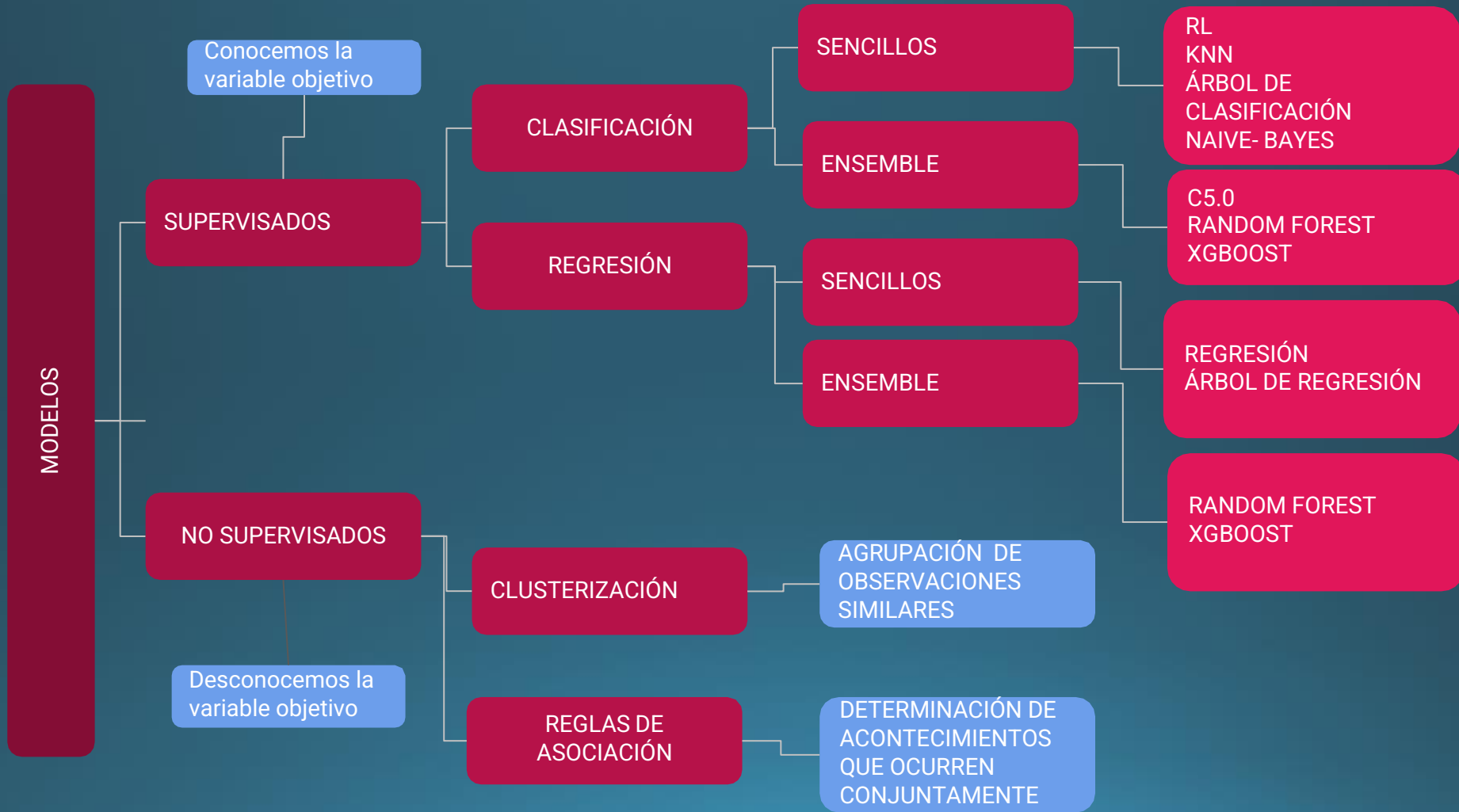
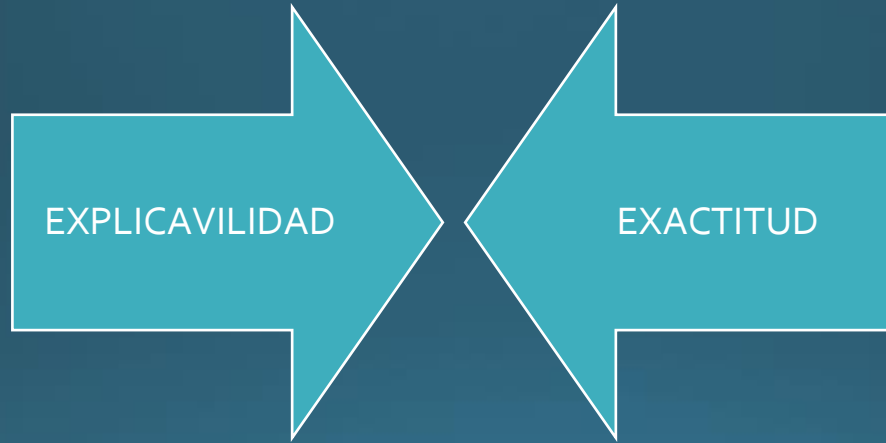


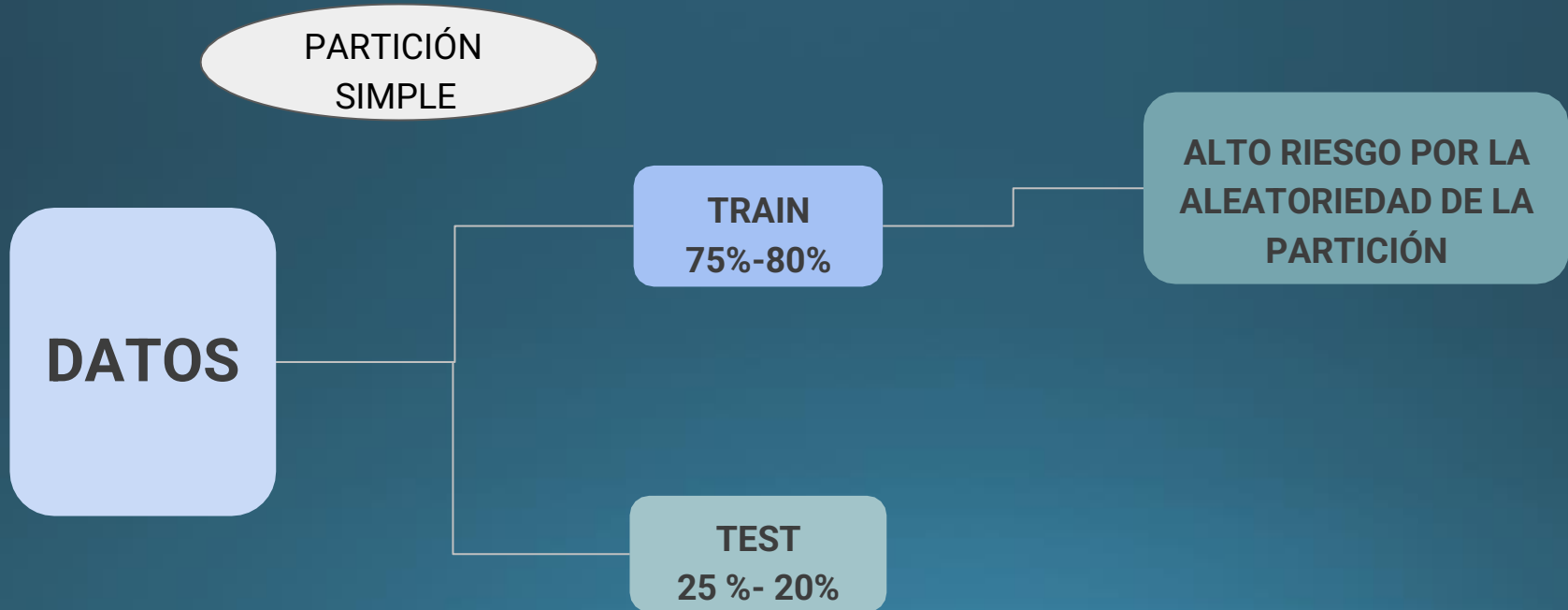
MODELOS



LA GRAN DISYUNTIVA



PARTICIONES



VALIDACIÓN
CRUZADA

1
2
3
4
5

1º PARTICIÓN
TRAIN=1,2,3,4
TEST=5

2º PARTICIÓN
TRAIN=1,2,3,5
TEST=4

3º PARTICIÓN
TRAIN=1,2,4,5
TEST=3

4º PARTICIÓN
TRAIN=1,3,4,5
TEST=2

5º PARTICIÓN
TRAIN=2,3,4,5
TEST=1

ELIMINACIÓN DEL
RIESGO ASOCIADO A
LA PARTICIÓN
SIMPLE

VALIDACIÓN MODELOS DE CLASIFICACIÓN

		Clase predicha	
		+	-
Clase real	+	TP True Positives	FN False Negatives Type II error
	-	FP False Positives Type I error	TN True Negatives

VALIDACIÓN MODELOS DE CLASIFICACIÓN

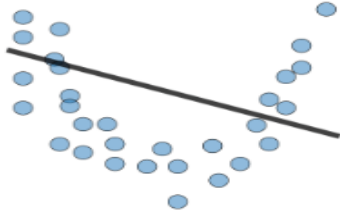
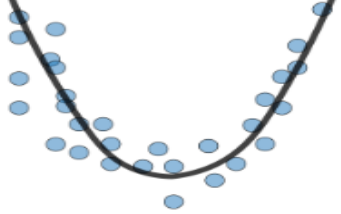
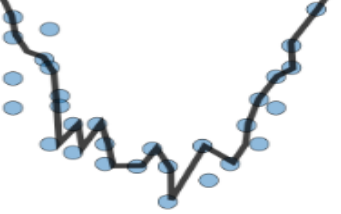
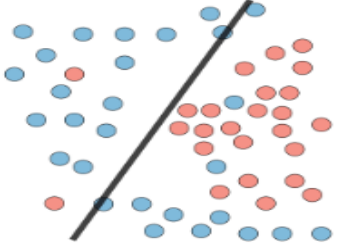
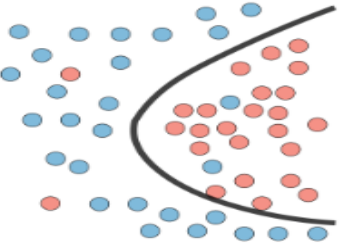
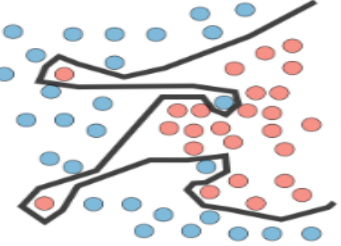
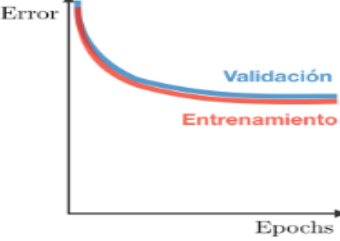
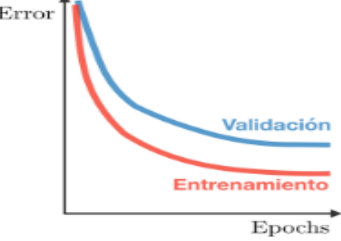
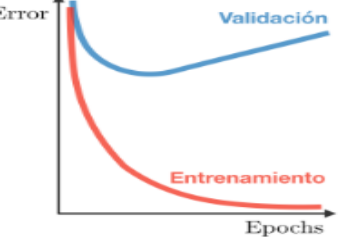
Métrica	Fórmula	Interpretación
Exactitud	$\frac{TP + TN}{TP + TN + FP + FN}$	Rendimiento general del modelo
Precisión	$\frac{TP}{TP + FP}$	Que tan precisas son las predicciones positivas
Exhaustividad Sensibilidad	$\frac{TP}{TP + FN}$	Cobertura de la muestra positiva real
Especificidad	$\frac{TN}{TN + FP}$	Cobertura de la muestra negativa real
F1 score	$\frac{2TP}{2TP + FP + FN}$	Métrica híbrida útil para clases desbalanceadas

VALIDACIÓN MODELOS DE REGRESIÓN

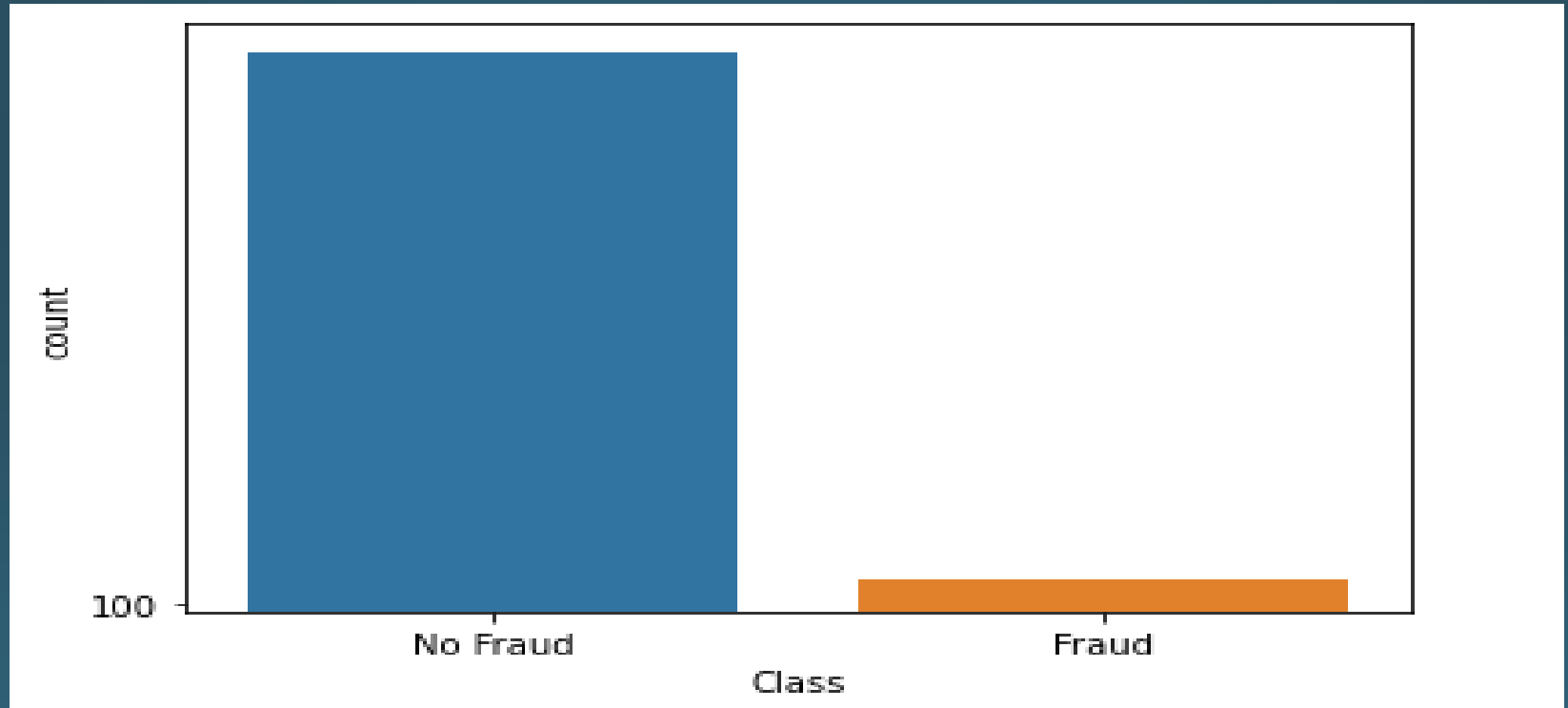
Suma total de cuadrados	Suma de cuadrados explicada	Suma residual de cuadrados
$SS_{\text{tot}} = \sum_{i=1}^m (y_i - \bar{y})^2$	$SS_{\text{reg}} = \sum_{i=1}^m (f(x_i) - \bar{y})^2$	$SS_{\text{res}} = \sum_{i=1}^m (y_i - f(x_i))^2$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

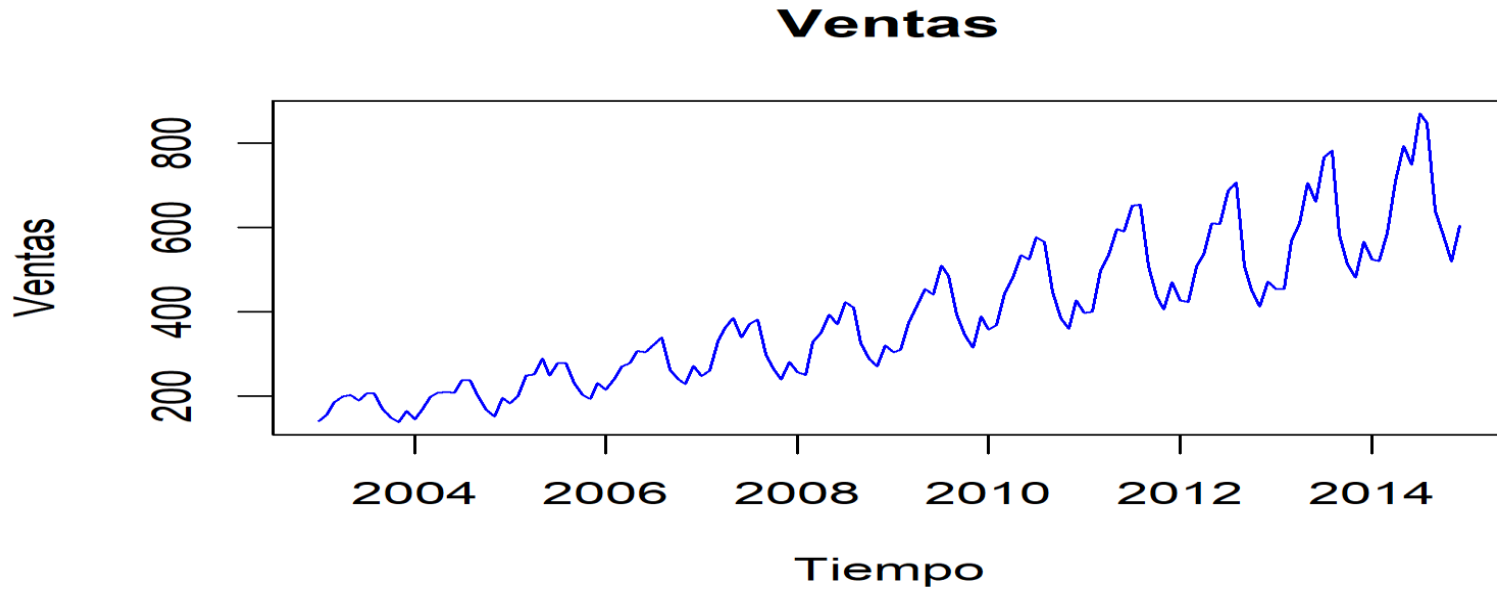
DIAGNÓSTICO

	Underfitting	Just right	Overfitting
Síntomas	<ul style="list-style-type: none">• Error de entrenamiento alto• Error de entrenamiento cercano al error de prueba• Sesgo alto	<ul style="list-style-type: none">• Error de entrenamiento ligeramente inferior al error de prueba	<ul style="list-style-type: none">• Error de entrenamiento muy bajo• Error de entrenamiento mucho más bajo que el error de prueba• Varianza alta
Ejemplo de regresión			
Ejemplo de clasificación			
Ejemplo de aprendizaje profundo			

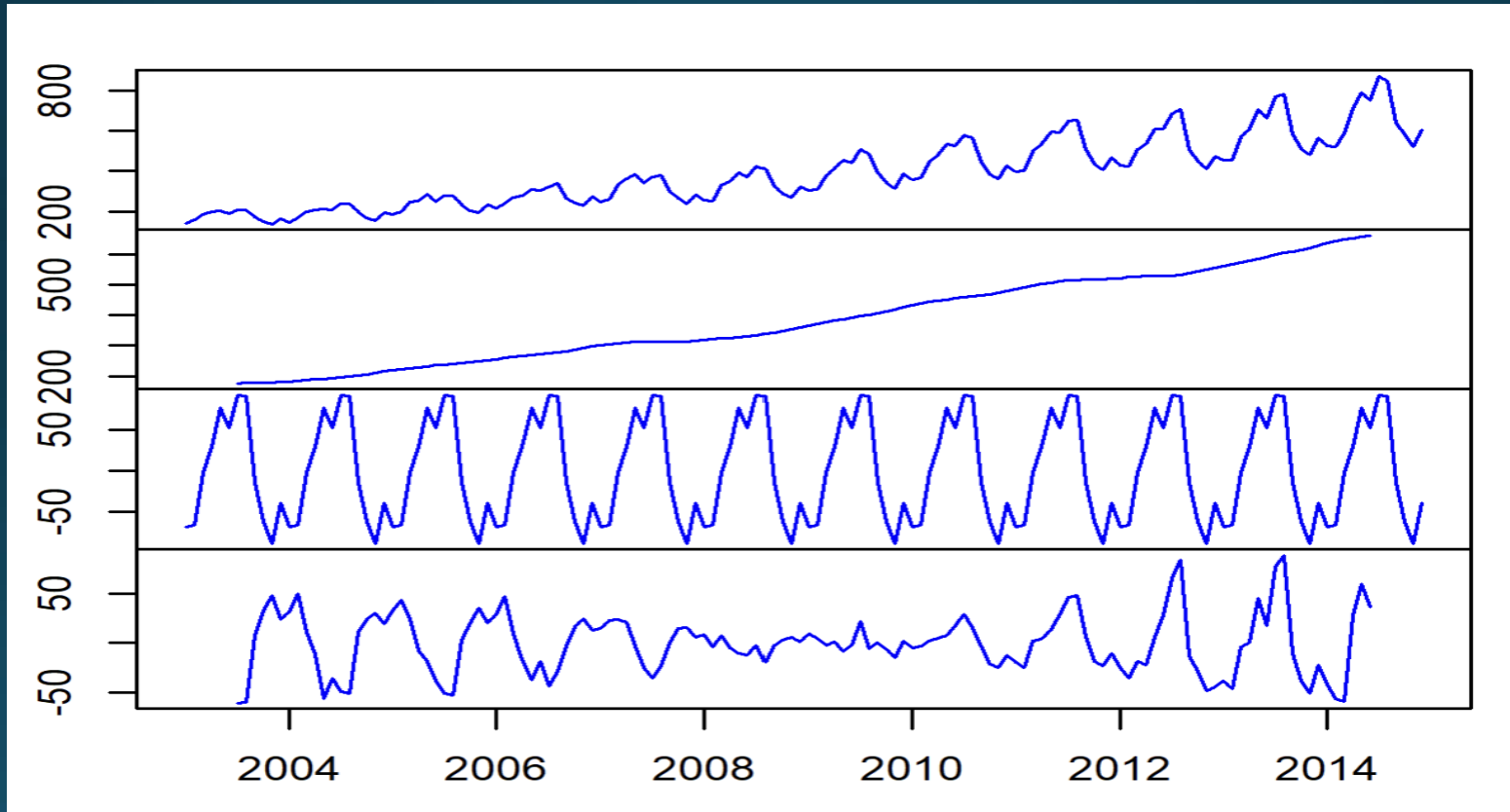
OTROS PROBLEMAS: IMBALANCEO



SERIES TEMPORALES

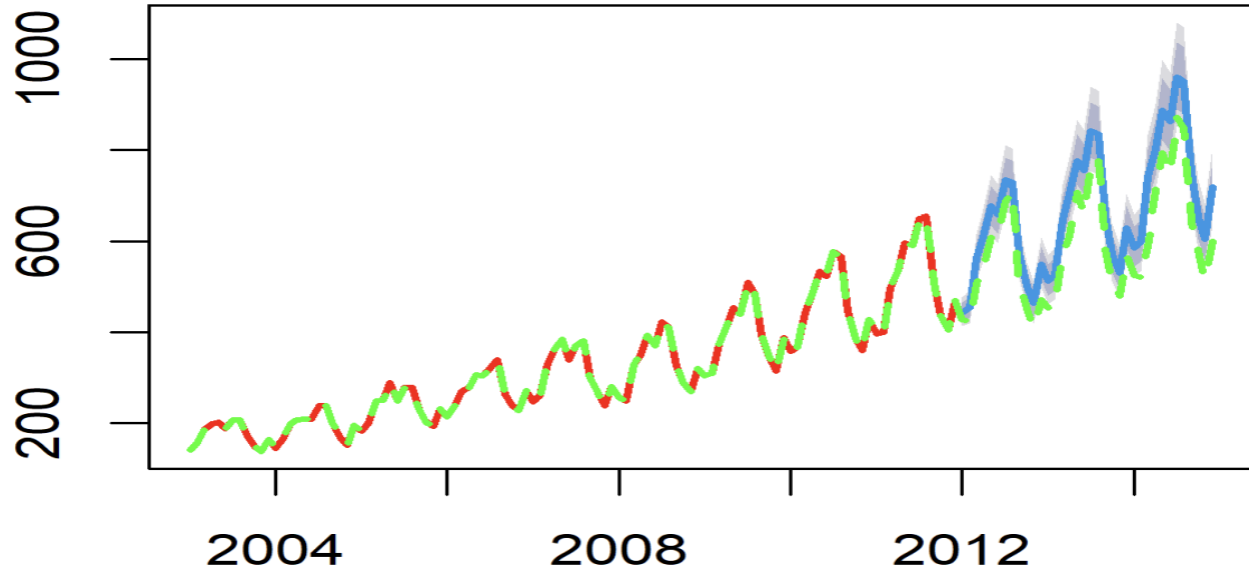


DESCOMPOSICIÓN DE LA SERIE



PREDICCIONES

Prediccion ARIMA



HOJA PROYECTOS BAIC

- <https://www.baic.eus/prensa/archivos/casos-uso.pdf>



CONTACTO

- Borja Balparda de Marco
- bbalparda@datavaluemanagement.es