



**DATA SCIENCE
& BIG DATA**
RESEARCH LAB



Workshop del 24-abril-2021

Predicción de Series Temporales

Prof. Dr. Gualberto Asencio Cortés

guaasecor@upo.es <http://datalab.upo.es/asencio>

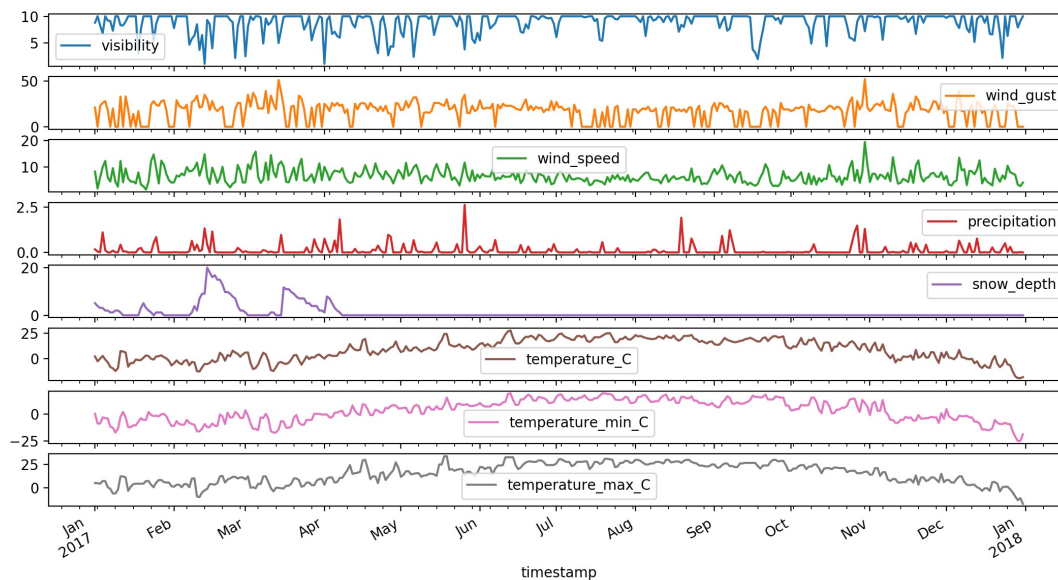
Contenidos

- ▶ ¿Qué es una serie temporal?
- ▶ ¿En qué consiste la predicción de series temporales?
- ▶ Metodología
 - ▶ Preprocesado de datos
 - ▶ Validación
 - ▶ Evaluación
 - ▶ Algoritmos de aprendizaje

¿Qué es una serie temporal?



- **Concepto de serie temporal**
 - Conjunto de observaciones ordenadas en el tiempo
- **Tipos de series temporales**
 - Series univariantes
 - Series multivariantes
- **Ejemplo de serie temporal multivariante meteorológica**



<https://towardsdatascience.com/prediction-task-with-multivariate-timeseries-and-var-model-47003f629f9>

¿En qué consiste la predicción de series temporales?

- **Predicción (*prediction, forecasting*)**
 - Estimación de valores futuros de una o varias variables objetivo (targets) de la serie
- **Horizonte de predicción (*prediction horizon*)**
 - Número de valores futuros a predecir
- **Intervalos de confianza**
 - Rango en el que se estima que se encontrará un valor futuro

■ Ejemplos:

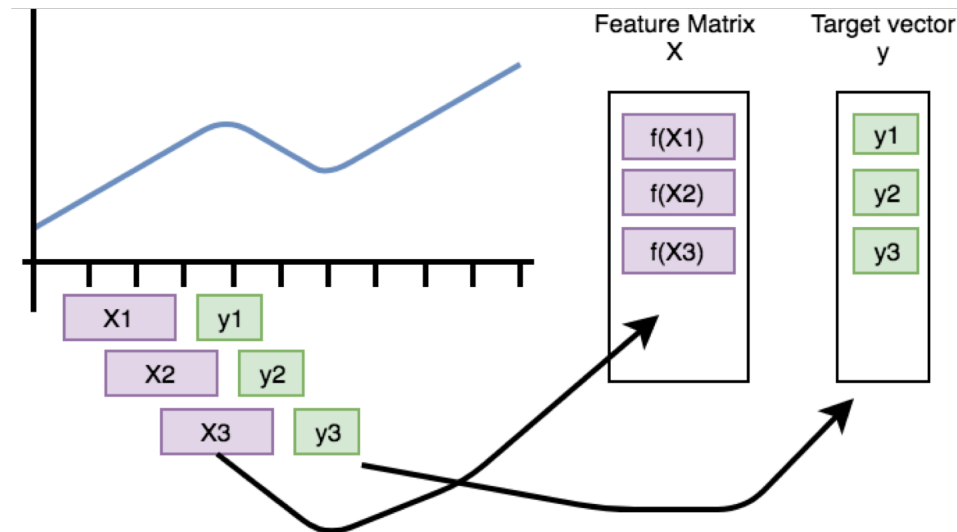
Forecasting future 10-year returns for the S&P 500
Actuals vs forecasts for different time series models
Look-ahead bias-free forecasts after the vertical red line



<https://www.r-bloggers.com/2020/05/forecasting-the-next-decade-in-the-stock-market-using-time-series-models/>

<https://github.com/KaroRonty/StockMarketLongTermForecast>

- **Transformación a conjunto de datos proposicional**
 - Transformación de la serie temporal a un conjunto de datos con atributos y clase, apto para aprendizaje automático
- **Ingeniería de atributos (*feature extraction and engineering*)**
 - Creación de variables en instantes pasados (*lag variables*)
 - Creación de variables en horizontes futuros (*ahead variables*)
 - Funciones y estadísticos (p.ej. media móvil) (*rolling statistics*)



- Ejemplo de transformación de serie temporal multivariante a conjunto de datos proposicional

SERIE TEMPORAL		
TIME	V1	V2
1	4	4
2	5	5
3	3	8
4	4	7
5	5	12
6	6	13
7	6	10
8	4	9
9	7	7
10	2	5
11	1	6
12	3	7



TRAIN SET							TARGETS			
V1_Lag-3	V1_Lag-2	V1_Lag-1	V2_Lag-3	V2_Lag-2	V2_Lag-1	V1_Ahead1	V1_Ahead2	V2_Ahead1	V2_Ahead2	
NA	NA	4	NA	NA	4	5	3	5	8	
NA	4	5	NA	4	5	3	4	8	7	
4	5	3	4	5	8	4	5	7	12	
5	3	4	5	8	7	5	6	12	13	
3	4	5	8	7	12	6	6	13	10	
				...						
4	7	2	9	7	5	1	3	6	7	

- Ejemplo de transformación de serie temporal multivariante a conjunto de datos proposicional

SERIE TEMPORAL		
TIME	V1	V2
1	4	4
2	5	5
3	3	8
4	4	7
5	5	12
6	6	13
7	6	10
8	4	9
9	7	7
10	2	5
11	1	6
12	3	7

TRAIN SET						TARGETS			
V1_Lag-3	V1_Lag-2	V1_Lag-1	V2_Lag-3	V2_Lag-2	V2_Lag-1	V1_Ahead1	V1_Ahead2	V2_Ahead1	V2_Ahead2
NA	NA	4	NA	NA	4	5	3	5	8
NA	4	5	NA	4	5	3	4	8	7
4	5	3	4	5	8	4	5	7	12
5	3	4	5	8	7	5	6	12	13
3	4	5	8	7	12	6	6	13	10
				...					
4	7	2	9	7	5	1	3	6	7

- Ejemplo de transformación de serie temporal multivariante a conjunto de datos proposicional

SERIE TEMPORAL		
TIME	V1	V2
1	4	4
2	5	5
3	3	8
4	4	7
5	5	12
6	6	13
7	6	10
8	4	9
9	7	7
10	2	5
11	1	6
12	3	7
13	3	6
14	7	7
15	5	8
16	3	5
17	45	4
18	2	6
19	3	9
20	12	12

TRAIN SET						TARGETS			
V1_Lag-3	V1_Lag-2	V1_Lag-1	V2_Lag-3	V2_Lag-2	V2_Lag-1	V1_Ahead1	V1_Ahead2	V2_Ahead1	V2_Ahead2
NA	NA	4	NA	NA	4	5	3	5	8
NA	4	5	NA	4	5	3	4	8	7
4	5	3	4	5	8	4	5	7	12
5	3	4	5	8	7	5	6	12	13
3	4	5	8	7	12	6	6	13	10
4	7	2	9	7	5	1	3	6	7

TEST SET						TARGETS			
V1_Lag-3	V1_Lag-2	V1_Lag-1	V2_Lag-3	V2_Lag-2	V2_Lag-1	V1_Ahead1	V1_Ahead2	V2_Ahead1	V2_Ahead2
2	1	3	5	6	7	PRED (3)	PRED (7)	PRED (6)	PRED (7)
1	3	3	6	7	6	PRED (7)	PRED (5)	PRED (7)	PRED (8)
3	45	2	5	4	6	PRED (3)	PRED (12)	PRED (9)	PRED (12)

- Ejemplo de transformación de serie temporal multivariante a conjunto de datos proposicional

SERIE TEMPORAL		
TIME	V1	V2
1	4	4
2	5	5
3	3	8
4	4	7
5	5	12
6	6	13
7	6	10
8	4	9
9	7	7
10	2	5
11	1	6
12	3	7

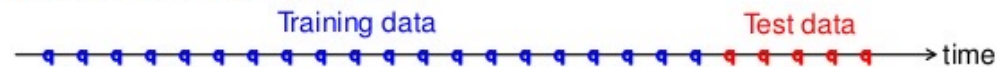
TRAIN SET						TARGETS			
V1_Lag-3	V1_Lag-2	V1_Lag-1	V2_Lag-3	V2_Lag-2	V2_Lag-1	V1_Ahead1	V1_Ahead2	V2_Ahead1	V2_Ahead2
NA	NA	4	NA	NA	4	5	3	5	8
NA	4	5	NA	4	5	3	4	8	7
4	5	3	4	5	8	4	5	7	12
5	3	4	5	8	7	5	6	12	13
3	4	5	8	7	12	6	6	13	10
4	7	2	9	7	5	1	3	6	7

13	3	6
14	7	7
15	5	8
16	3	5
17	45	4
18	2	6
19	3	9
20	12	12

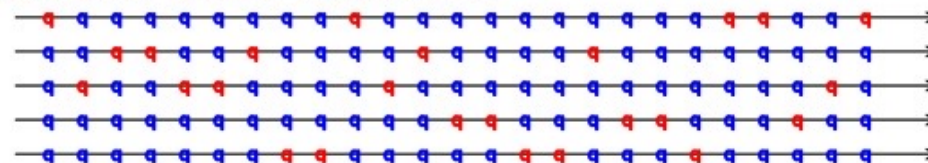
TEST SET						TARGETS			
V1_Lag-3	V1_Lag-2	V1_Lag-1	V2_Lag-3	V2_Lag-2	V2_Lag-1	V1_Ahead1	V1_Ahead2	V2_Ahead1	V2_Ahead2
2	1	3	5	6	7	PRED (3)	PRED (7)	PRED (6)	PRED (7)
1	3	3	6	7	6	PRED (7)	PRED (5)	PRED (7)	PRED (8)
3	45	2	5	4	6	PRED (3)	PRED (12)	PRED (9)	PRED (12)

- **Proceso:** división de los datos en conjuntos de training y test
- **Comparación con técnicas de validación tradicionales**

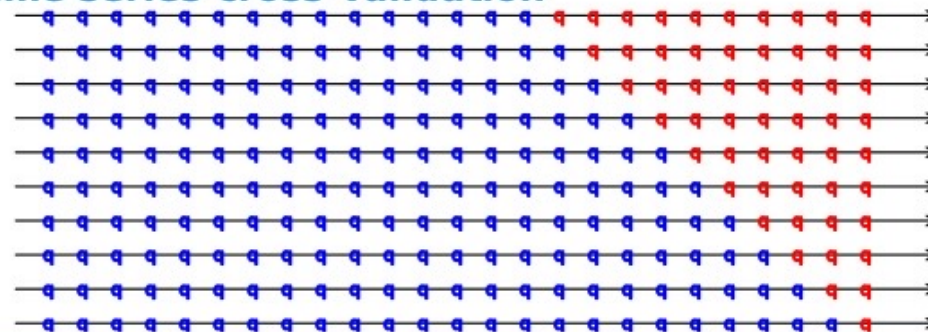
Traditional evaluation



Standard cross-validation

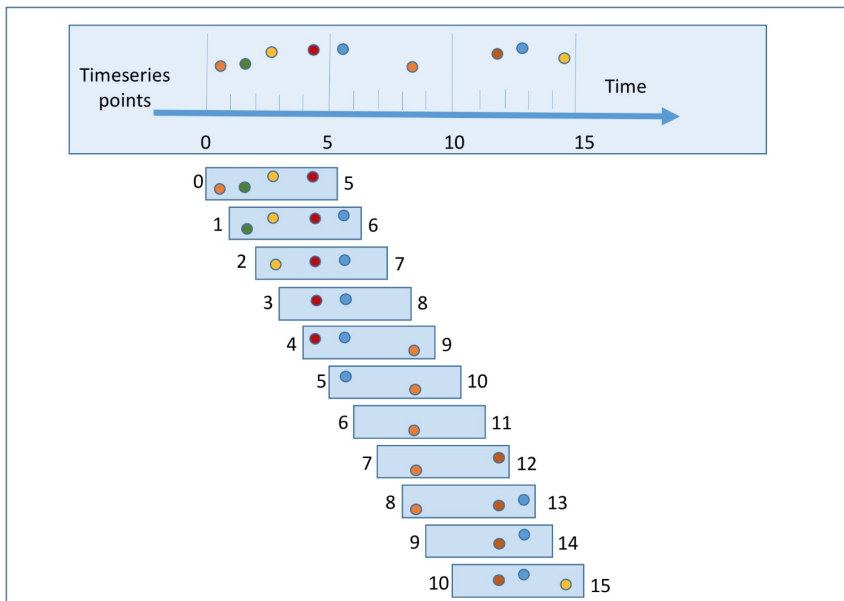


Time series cross-validation

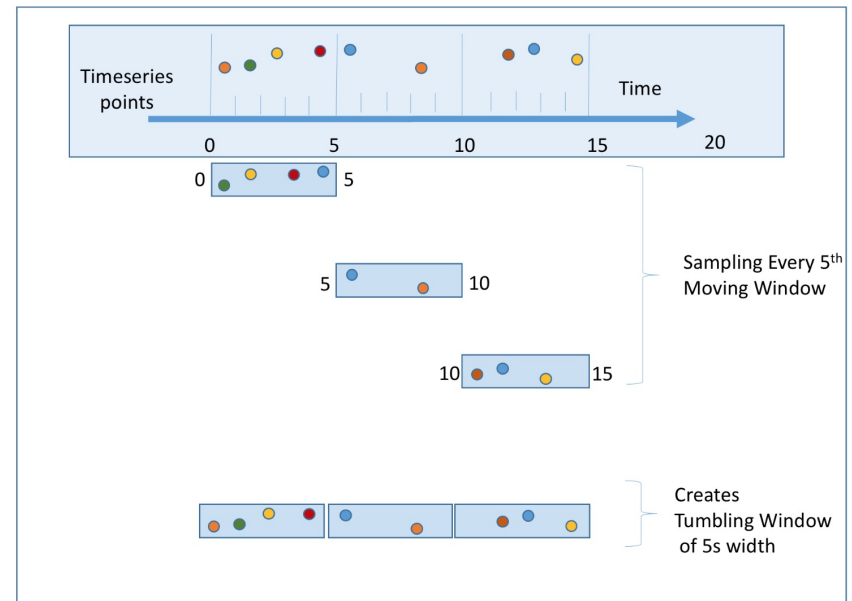


■ Dos estrategias de desplazamiento con ventana deslizante

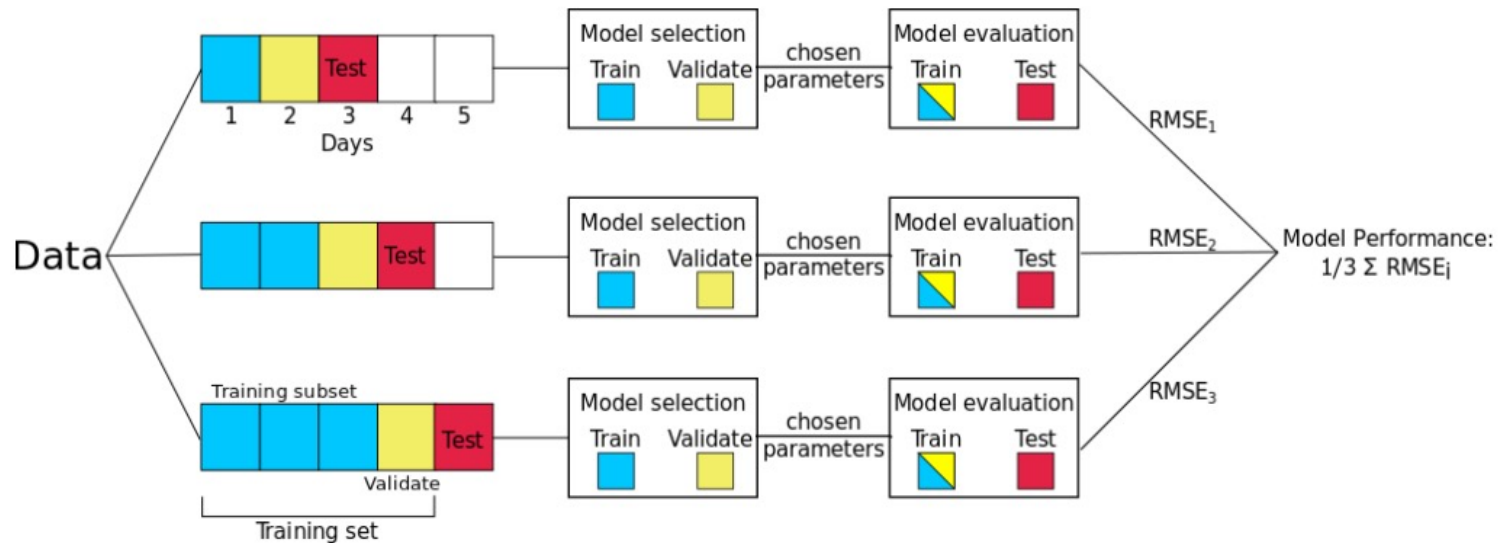
■ Con solapamiento



■ Sin solapamiento



Optimización de hiperparámetros mediante validación anidada



- **Métricas de evaluación en regresión**

- MAE, RMSE, MSE, MAPE, R², etc.

$$MAE = \frac{1}{n} \sum_{h=1}^n |y_h - \hat{y}_h| \quad RMSE = \sqrt{\frac{1}{n} \sum_{h=1}^n |y_h - \hat{y}_h|^2} \quad MAPE(\%) = \frac{100}{n} \sum_{h=1}^n \frac{|y_h - \hat{y}_h|}{y_h}$$

- **Ejemplos:**

Table 2

Summary of results obtained in terms of MAE, MAPE and RMSE.

Algorithm	MAE	MAPE	RMSE
ARIMA	146.56	13.40	142.37
CART	137.59	10.97	177.34
ANN	106.44	9.78	141.20
GBM	69.71	5.79	93.75
PSF	95.85	7.94	123.92
7-PSF	68.33	5.63	93.18
bigPSF	57.15	4.70	61.23

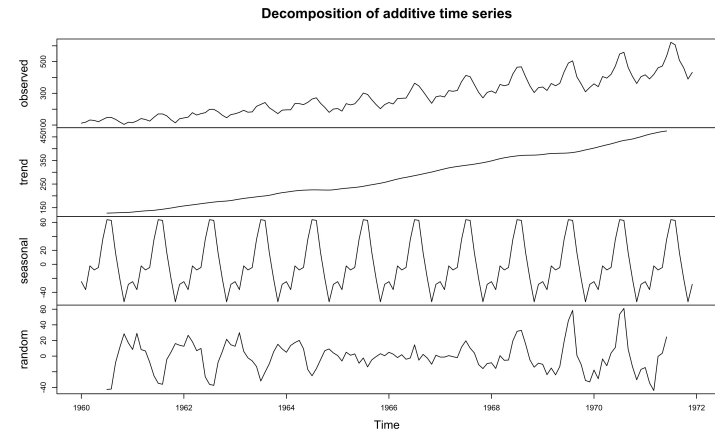
<https://doi.org/10.1016/j.ins.2020.06.014>

Table 4. Summary of the achieved results in terms of mean absolute error (MAE) (MW), mean absolute percentage error (MAPE) (%), and Root Mean Squared Error (RMSE).

	K	funPSF	7-funPSF	ANN	ARIMA
MAE		95.85	68.33	106.44	146.56
MAPE		7.94	5.63	9.78	13.40
RMSE		123.92	93.18	141.20	142.37

<http://doi.org/10.3390/en12010094>

- **Algoritmos de regresión (generalmente)**
 - Regresión univariante (*single output*)
 - Regresión multivariante (*multi output*)
- **Algoritmos estadísticos**
 - AR/X (VAR), ARMA/X, ARIMA/X,
 - GARCH, ETS, HoltWinters, etc.
- **Regresores clásicos**
 - Base: LR, IBk, M5P, GaussianProcesses, MLP, SMOreg, etc.
 - Ensembles: Random Forest (bagging), Gradient Boosting Machines o XGBoost (boosting), Stacking, etc.
- **Deep Learning**
 - RNN, LSTM, etc. (mecanismo attention)
- **Transfer Learning** (datos insuficientes en dominio objetivo)
- **Online Learning** (aprendizaje incremental en entornos data streaming)



MUCHAS GRACIAS ¿PREGUNTAS?



**DATA SCIENCE
& BIG DATA**
RESEARCH LAB



Prof. Dr. Gualberto Asencio Cortés

guaasecor@upo.es <http://datalab.upo.es/asencio>