

Modelo de Pronóstico de Series Temporales

Proyecto: Themis
A.I. Prediction Framework
Modelo de Pronóstico de Series Temporales

REGRESOR AL FUTURO

apsl

XGBOOST



Jaume Perelló Perelló



Juan Carlos Gonzalez-Avella

Índice

Introducción: Themis

Core: Delfos

Casos de Estudio

Funcionamiento de Themis



Licurgo, el gran legislador de Esparta, consultando a la pitia de Delfos. Según la tradición, ella le dio el corpus de leyes por el que se rigió Esparta. Eugène Delacroix (1835-45), University of Michigan Museum of Art.

Introducción



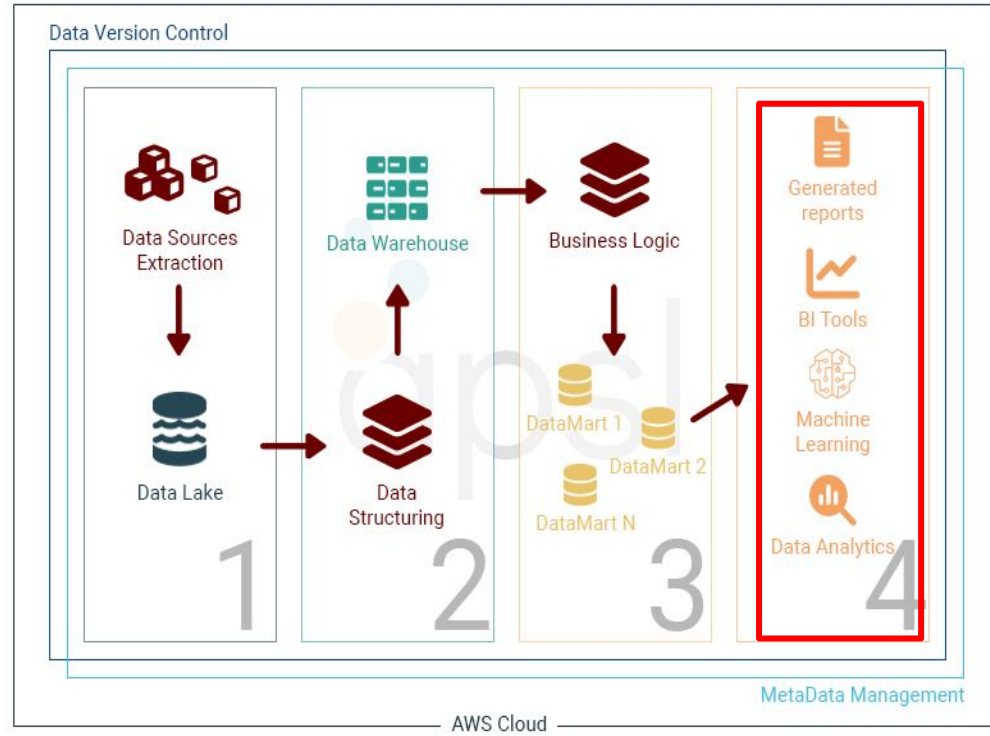
XGBOOST

- *Taylor made service* para empresas que implementan **CIF** en entornos **CLOUD**.

- Combinar diferentes fuentes de datos
- Establecer un Data Lake
- Transformar datos y contenerlos en Data WareHouses
- Distribuirlo en Data Marts autónomos listos para ser empleados

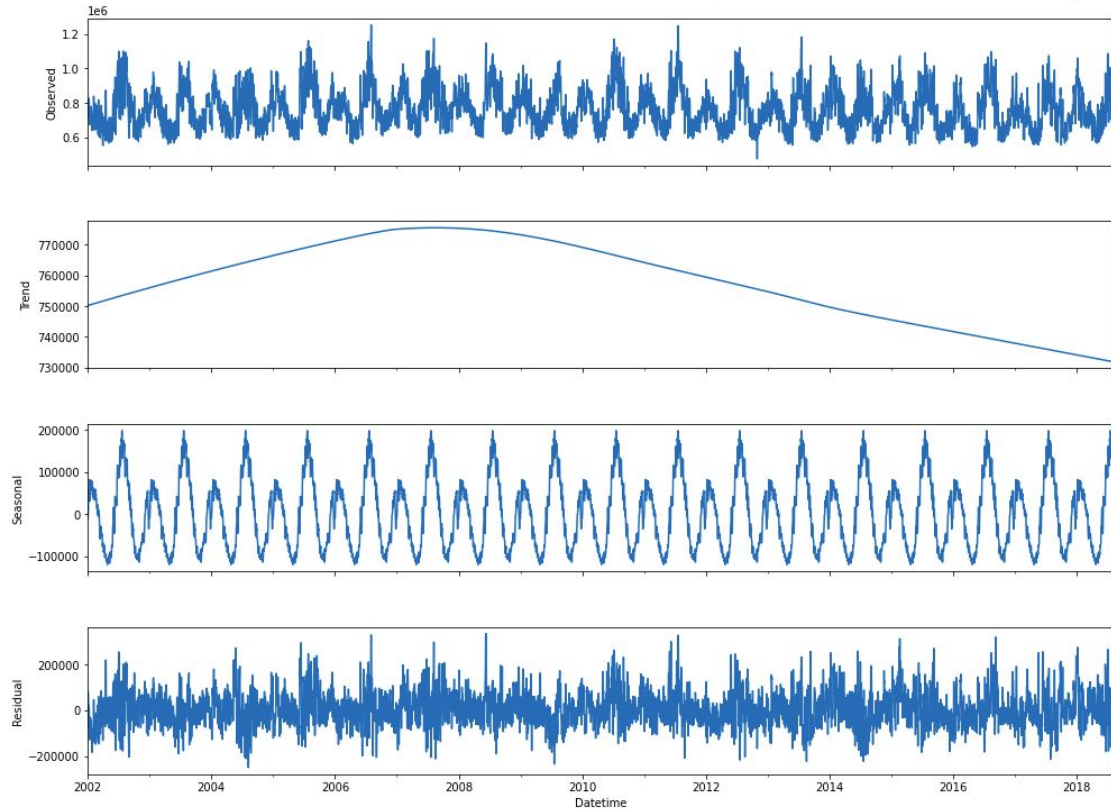
- Explotación de datos:
 - Presentación de insights - BI
 - ML - Data Analytics

DaaS - Data as a Service

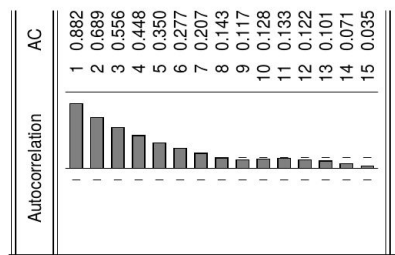
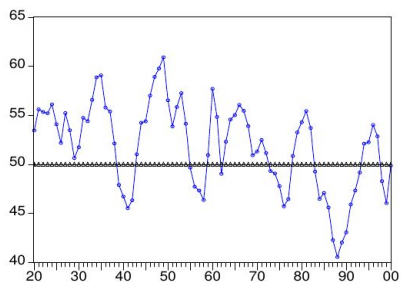


- Las series temporales son una colección de puntos de datos dependientes del tiempo. Esto significa que cada punto de datos se asigna a una marca de tiempo específica.
- en orden cronológico y en intervalos de tiempo regulares (continuos o discretos).
- Las series temporales dependen del tiempo y muestran características específicas, tales como:
 - **Tendencia**
 - **Estacionalidad**
 - **Ruido(Blanco, de colores, otros)**

$$Y_t = PS_t + a_t \quad \longrightarrow \quad PS_t = f(Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots)$$



- AR (Autoregressive model)
- ARMA (Autoregressive-moving average model)
- ARIMA (Autoregressive-integrated moving average)
- SARIMA (Seasonal ARIMA)
- ARCH (Autoregressive conditional heteroskedasticity)
- GARCH (Generalized Autoregressive Conditional Heteroskedasticity)
- SV (Volatility Forecasting)
- Vectorial ARMA (VAR(p), VMA(q), VARMA(p, q))
- Markovian Processes
- Modelos Basados en Agentes



Notas importantes al implementar modelos AR

- Algoritmos muy complejos
- Requiere análisis de muchos correlogramas para ACF y ACPF
- Definir que modelo emplear
- Metodos que en general no consideran influencias externas (incluidas en la componente de ruido)

Boosting es un enfoque de **Machine Learning** basado en la idea de crear una regla de predicción altamente precisa combinando muchas reglas relativamente débiles e imprecisas.

El **Gradient boosting** implica tres elementos:

- Una **función de pérdida** a optimizar .
- Un *algoritmo de aprendizaje débil* para hacer las predicciones.
- Un modelo aditivo para añadir los *algoritmos de aprendizaje débiles* que minimizan la **función de pérdida**.

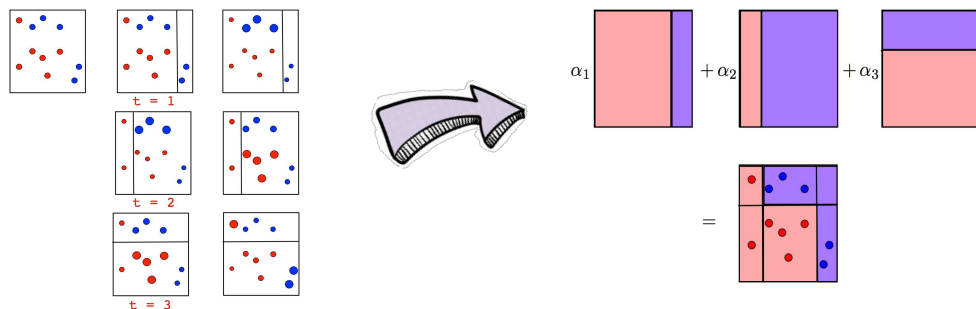
XGBoost es una implementación de **árboles de decisión** con **Gradient boosting** diseñada para minimizar la velocidad de ejecución y maximizar el rendimiento.

Modelo matemático regresión (tipo ARIMA)

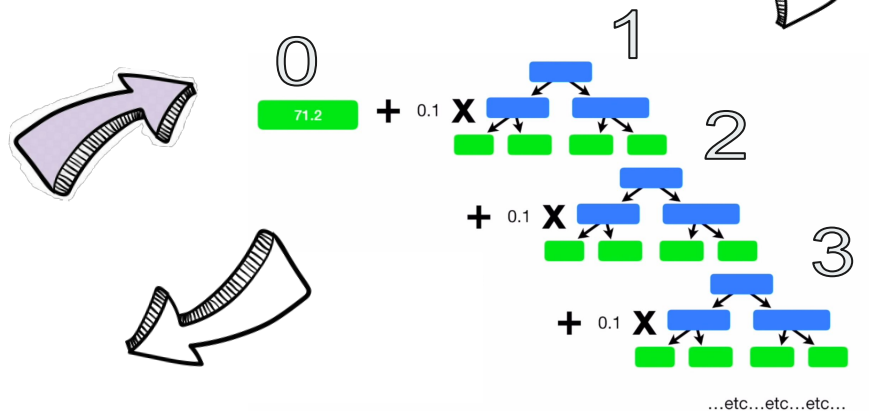
$$Y_t = -(\Delta^d Y_t - Y_t) + \phi_0 + \sum_{i=1}^p \phi_i \Delta^d Y_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

$$\Delta Y_t = Y_t - Y_{t-1}$$

Ilustración de la suma de algoritmos de *aprendizajes débiles*



Árboles de decisión con Gradient Boosting



Modelo matemático regresión (tipo ARIMA)

$$Y_t = -(\Delta^d Y_t - Y_t) + \phi_0 + \sum_{i=1}^p \phi_i \Delta^d Y_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

$$\Delta Y_t = Y_t - Y_{t-1}$$

Boosting es un enfoque de **Machine Learning** basado en la idea de crear una regla de predicción altamente precisa combinando muchas reglas relativamente débiles e imprecisas.

El **Gradient boosting** implica tres elementos:

- Una **función de pérdida** a optimizar .
- Un *algoritmo de aprendizaje débil* para hacer las predicciones.
- Un modelo aditivo para añadir los *algoritmos de aprendizaje débiles* que minimizan la **función de pérdida**.

XGBoost es una implementación de **árboles de decisión** con **Gradient boosting** diseñada para minimizar la velocidad de ejecución y maximizar el rendimiento.

Suma de algoritmos de *aprendizajes débiles*

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^n \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)),$$

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) - \gamma \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i))),$$

Input: training set $\{(x_i, y_i)\}_{i=1}^n$, a differentiable loss function $L(y, F(x))$, number of iterations M .

Algorithm:

1. Initialize model with a constant value:

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma).$$

2. For $m = 1$ to M :

1. Compute so-called *pseudo-residuals*:

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad \text{for } i = 1, \dots, n.$$

2. Fit a base learner (or weak learner, e.g. tree) $h_m(x)$ to pseudo-residuals, i.e. train it using the training set $\{(x_i, r_{im})\}_{i=1}^n$.
3. Compute multiplier γ_m by solving the following **one-dimensional optimization** problem:

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)).$$

4. Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

3. Output $F_M(x)$.

Core: Delfos



XGBOOST



Los conjuntos de datos concatenados alimentan el modelo tanto en la fase de entrenamiento (Ajuste) como en la fase de pronóstico (Predecir).

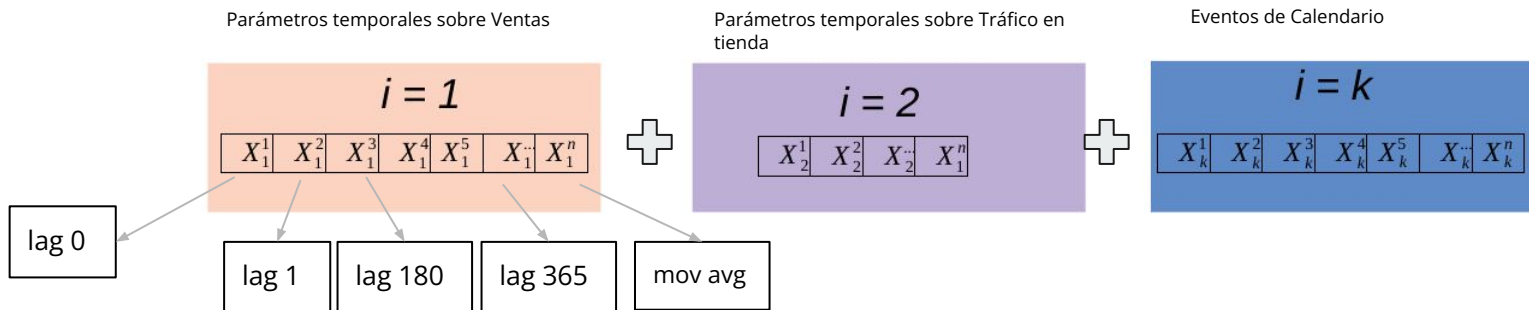
1. Una matriz conteniendo los diferentes parámetros(k) debe ser provista en la entrada de datos.
2. La concatenación de datos del marco de trabajo permite una estructuración modular de los pronósticos, permitiendo la flexibilidad necesaria para incluir o eliminar variables al modelo.

$$\text{concat} \left(\sum_{i=1}^k \sum_{j=1}^n \vec{X}_i^j \right)$$

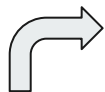
Delfos Forecasting

Aplicación que permite realizar pronósticos de comportamiento futuro en base al análisis de datos históricos, referencias autorregresivas y otros parámetros externos que tienen impacto en la variable objetivo dinámica (ej.: Ventas, Tráfico de tiendas minoristas)

Concatenación de Datos en Módulos



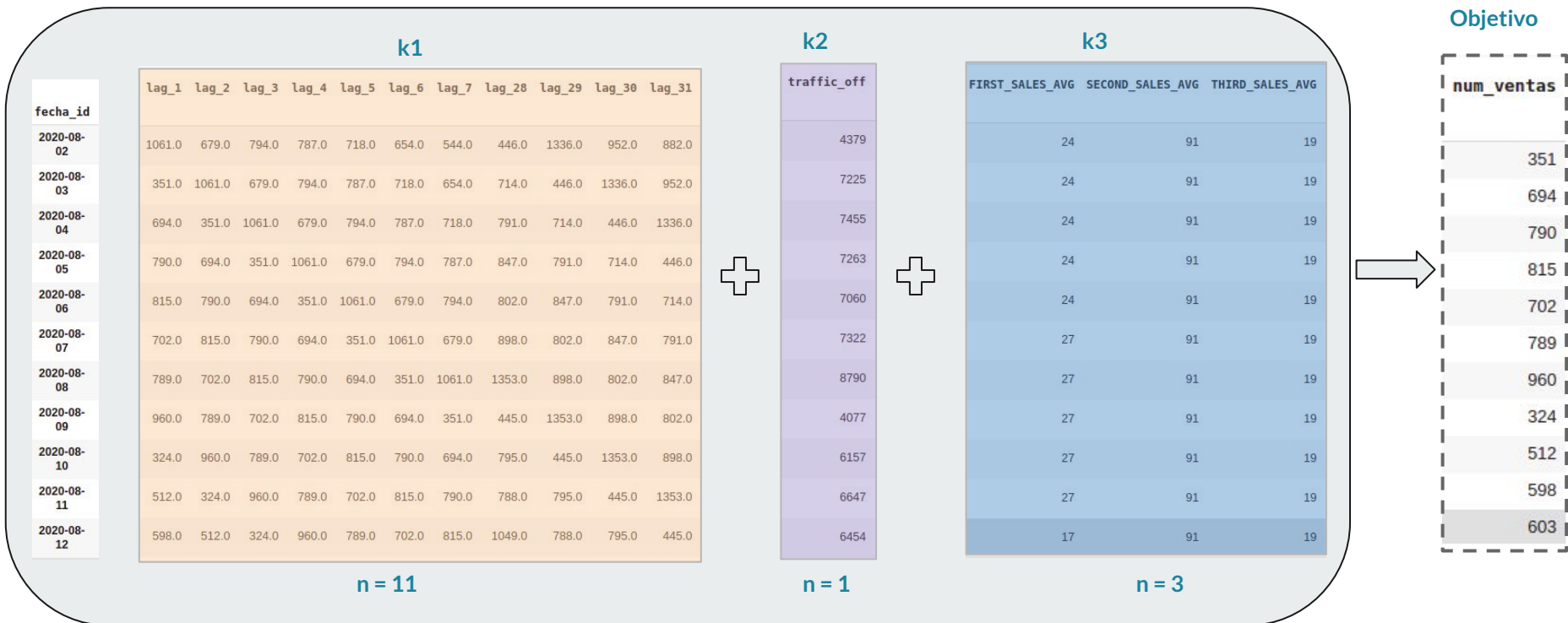
- k1 | Parámetros Autorregresivos
- k2 | Variable Temporal adicional
- k3 | Calendario de eventos *know in advance*



$$Prevision.fit(\text{concat}(\sum_{i=1}^k \sum_{j=1}^n \vec{X}_i^j))$$



Fase de Entrenamiento



Funcionamiento de la APP

- Interfaz
- Núcleo
- Visualización

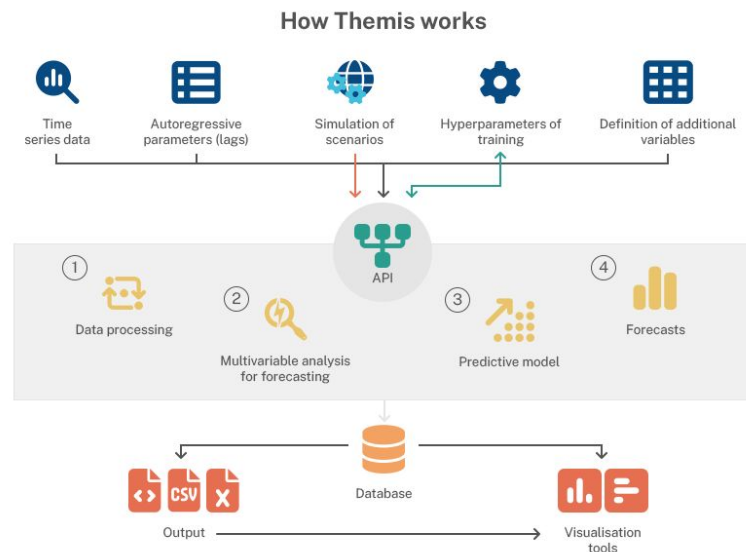
XGBOOST

1. Serie Temporal de Estudio
2. Requerimientos de la API para el pronóstico de series temporales
 - a. Data Mart
 - b. Parámetros Autorregresivos
 - c. **Pronósticos independientes**
 - d. Simulación de escenarios
 - e. Campos adicionales
 - f. Hiperparámetros de entrenamiento
3. Resultados
 - a. Archivos conteniendo las predicciones
 - b. métricas de evaluación estadística
4. Visualizaciones
 - a. Soporte Gráfico
 - b. Dashboard en herramienta de visualización

Requerimientos

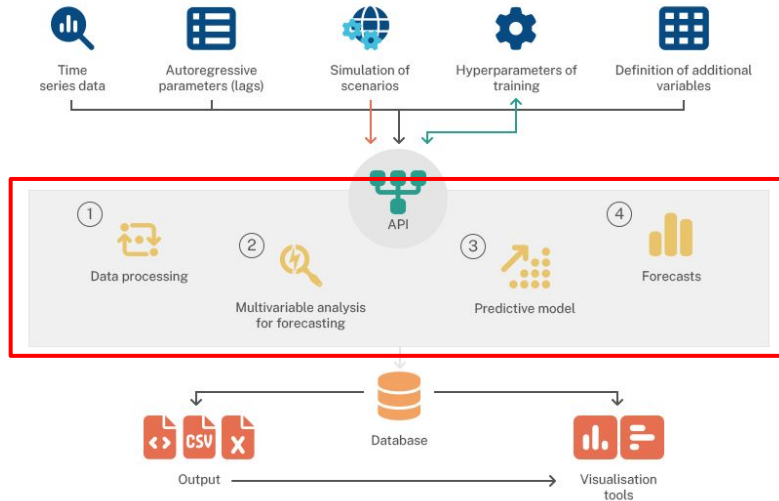
Format +
Pron. +
Eval.

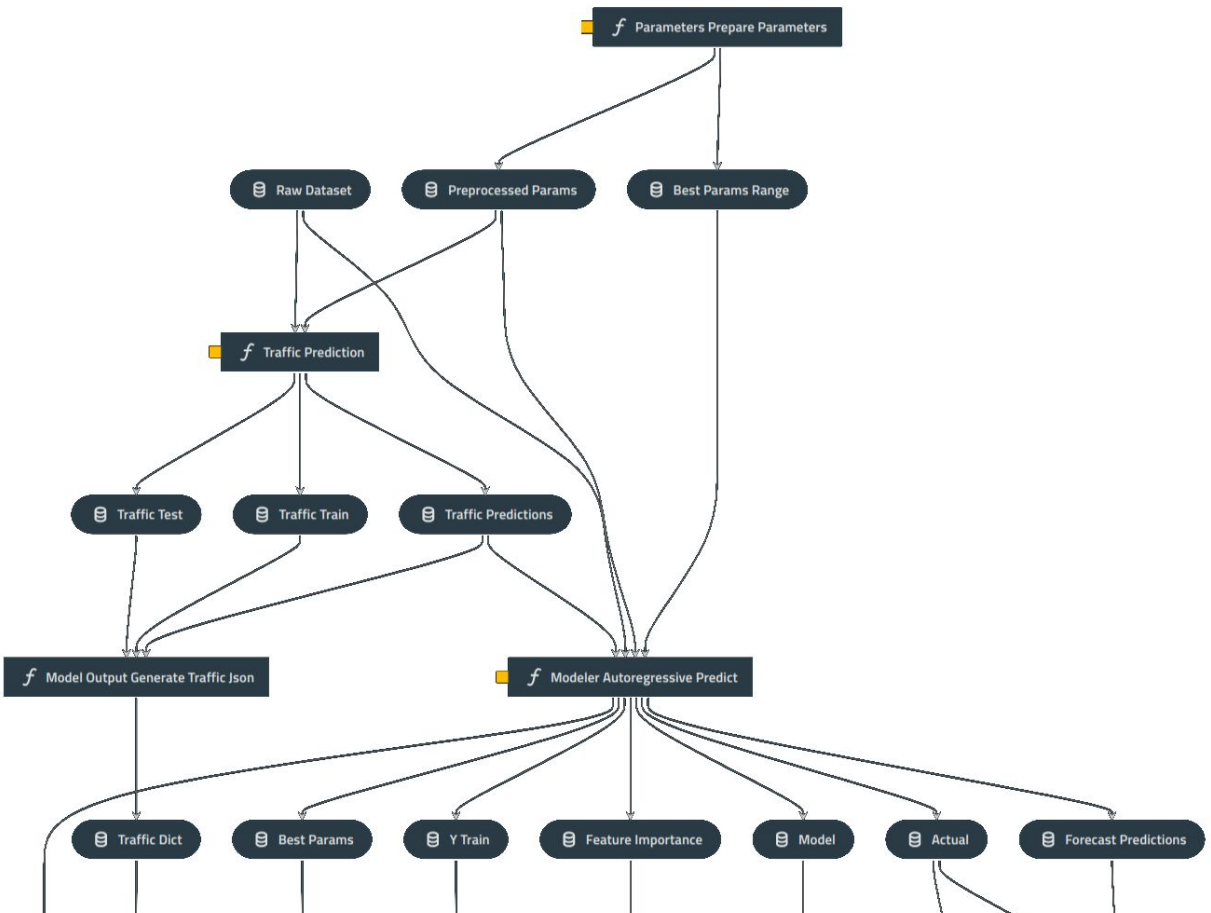
Tablas
Resultados



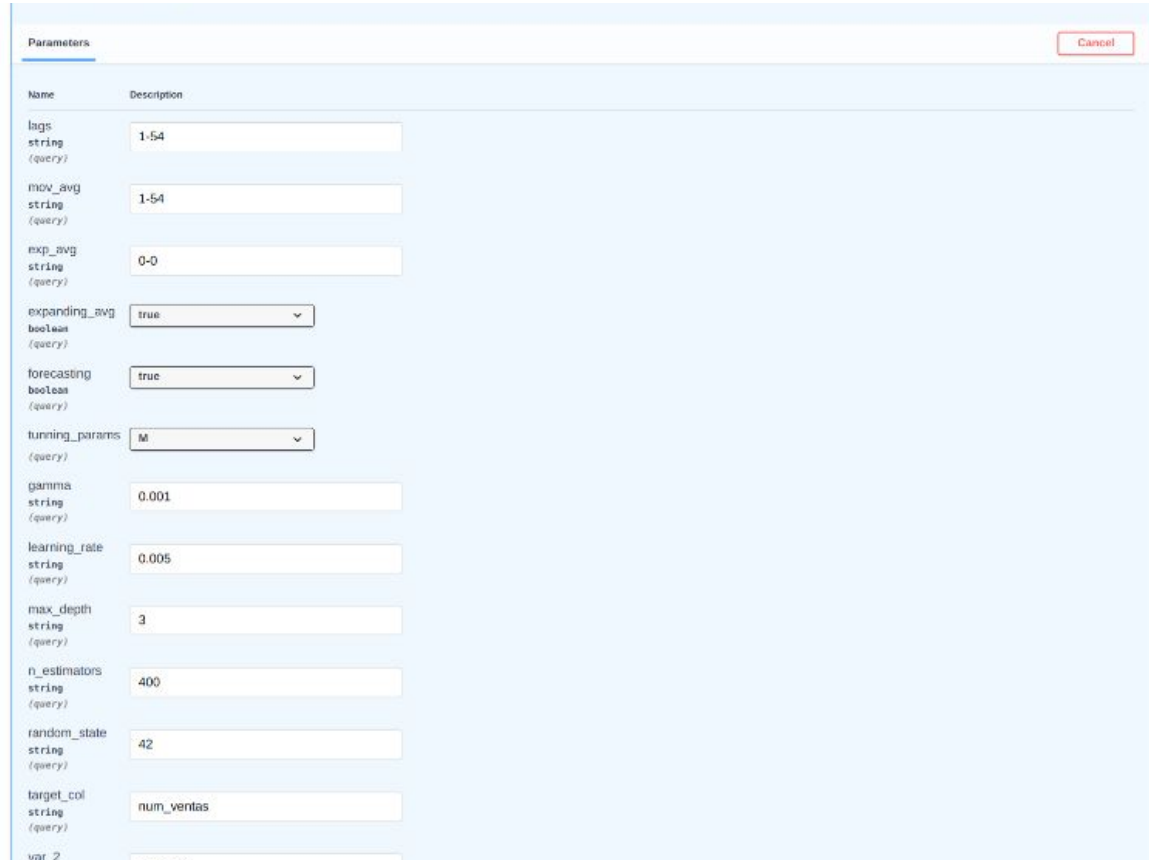
- Kedro es una biblioteca Python de código abierto, mantenida por QuantumBlack, que se enfoca en crear Pipelines de MLOps
- Ayuda a estructurar el código de manera que es fácil seguir el proceso que sigue el flujo de nuestros proyectos. Y pone estos procesos en producción siguiendo las buenas prácticas de ingeniería de software

How Themis works





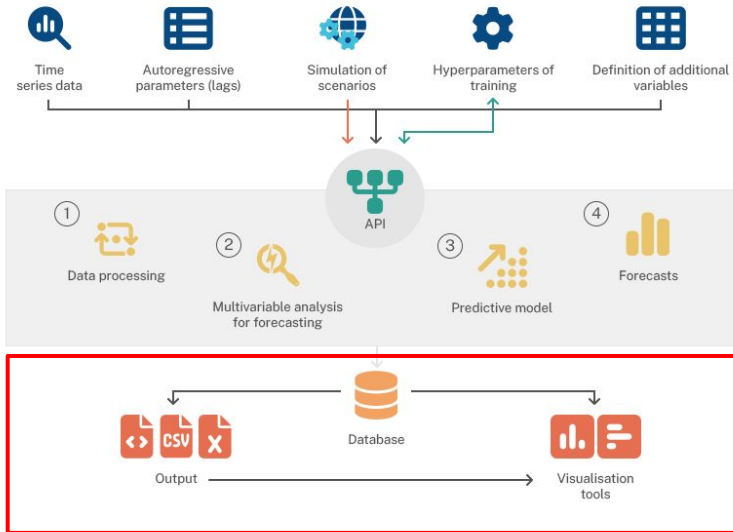
- Permite la validación de parámetros de entrada y testear nuestras queries usando su interfaz de testeo.
- Podemos usar la llamada de la api para automatizar forecasting periodico de los nuevos datos de nuestras series temporales



The screenshot shows a web interface titled "Parameters" with a "Cancel" button in the top right corner. The interface displays a list of parameters for a query, each with a "Name" and a "Description". The parameters and their values are as follows:

Name	Description
lags	1-54
mov_avg	1-54
exp_avg	0-0
expanding_avg	true
forecasting	true
tunning_params	M
gamma	0.001
learning_rate	0.005
max_depth	3
n_estimators	400
random_state	42
target_col	num_ventas
var_2	

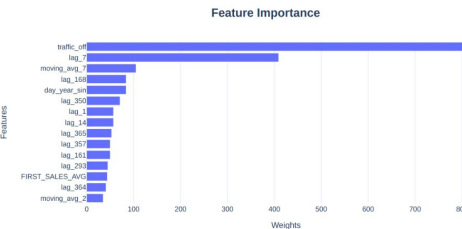
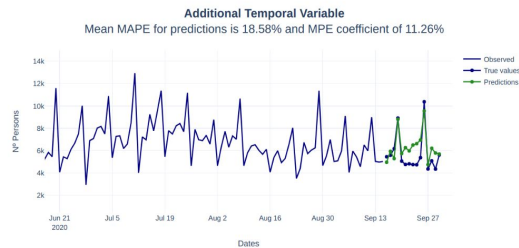
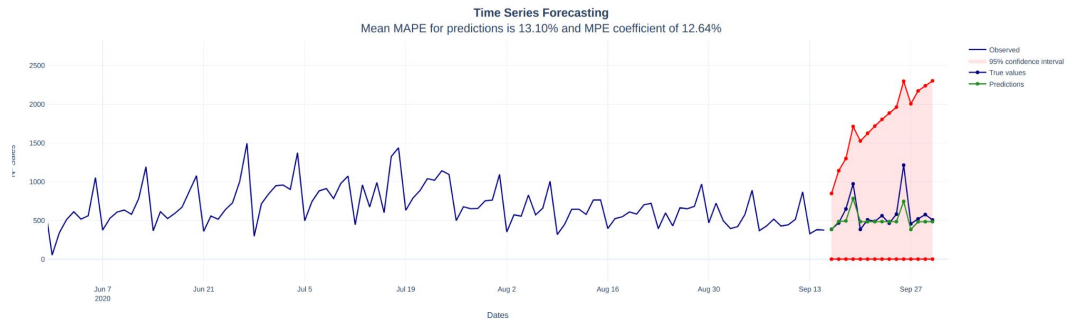
How Themis works



Time Series Forecasting Model

An APSL Application

Drag and Drop or Select a File



- El núcleo sigue siendo Kedro
- La interfaz para la introducción de datos y parámetros es similar a la de Fastapi
- Vemos los resultados de la predicción automáticamente
- Todos los datos y parámetros del experimento se guardan.
- Se puede acceder a predicciones anteriores y clonar el experimento refinando los parámetros o usando nuevos datos

Themis experiments

New Experiment

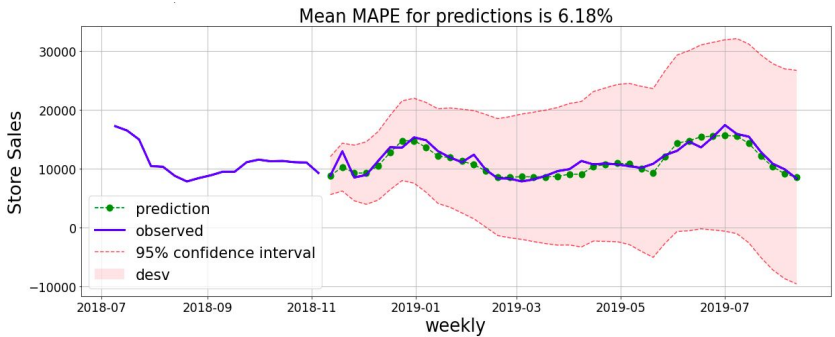
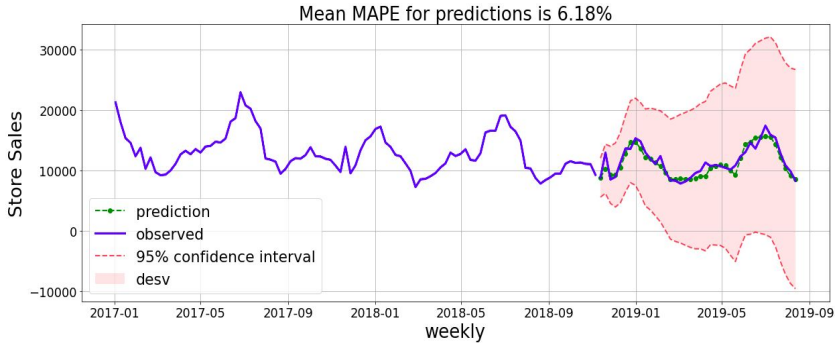
- Feb. 4, 2022, 9:50 a.m. 79e53a5 Experiment name |Graph |Json
- Feb. 4, 2022, 10:10 a.m. 3eb0250 Experiment name |Graph |Json
- July 19, 2022, 3:41 p.m. 73b6b7c Experiment name |Graph |Json
- July 19, 2022, 4:03 p.m. 51be23f Experiment name |Graph |Json
- July 19, 2022, 4:05 p.m. add05cd Experiment name |Graph |Json
- Sept. 12, 2022, 2:56 p.m. 9c9c3c3 Experiment name |Graph |Json
- Sept. 12, 2022, 3:09 p.m. afc9d74 Experiment name |Graph |Json
- Sept. 12, 2022, 3:14 p.m. d4aa949 Experiment name |Graph |Json
- Sept. 12, 2022, 3:15 p.m. f51e666 Experiment name |Graph |Json
- Sept. 12, 2022, 3:17 p.m. 3e7d9d1 Nuclear year |Graph |Json
- Sept. 12, 2022, 3:18 p.m. 7f28f8a Nuclear year |Graph |Json
- Sept. 12, 2022, 3:18 p.m. c6c868c Nuclear year |Graph |Json
- Sept. 12, 2022, 3:20 p.m. 4952c55 Experiment name |Graph |Json
- Sept. 12, 2022, 3:20 p.m. d7c8884 Experiment name |Graph |Json
- Sept. 12, 2022, 3:21 p.m. 342dadf Experiment name |Graph |Json
- Sept. 12, 2022, 3:23 p.m. d45ee2d Experiment name |Graph |Json
- Sept. 12, 2022, 3:23 p.m. a2e4e53 Experiment name |Graph |Json
- Sept. 12, 2022, 3:32 p.m. 354ac5d Experiment name |Graph |Json
- Sept. 12, 2022, 3:35 p.m. 9bb4980 Experiment name |Graph |Json
- Sept. 12, 2022, 3:39 p.m. 7b58413 Experiment name |Graph |Json
- Sept. 12, 2022, 3:39 p.m. 36b9bf7 Experiment name |Graph |Json
- Sept. 12, 2022, 3:41 p.m. 65416ef Experiment name |Graph |Json

N splits:

name	1.146087030810716
image	0.080222707400719
urlname	0.071070400007076
url_id	157.7024807804101
url_name	0.080222707400719
url_id	0.080222707400719

Casos de estudio

XGBOOST



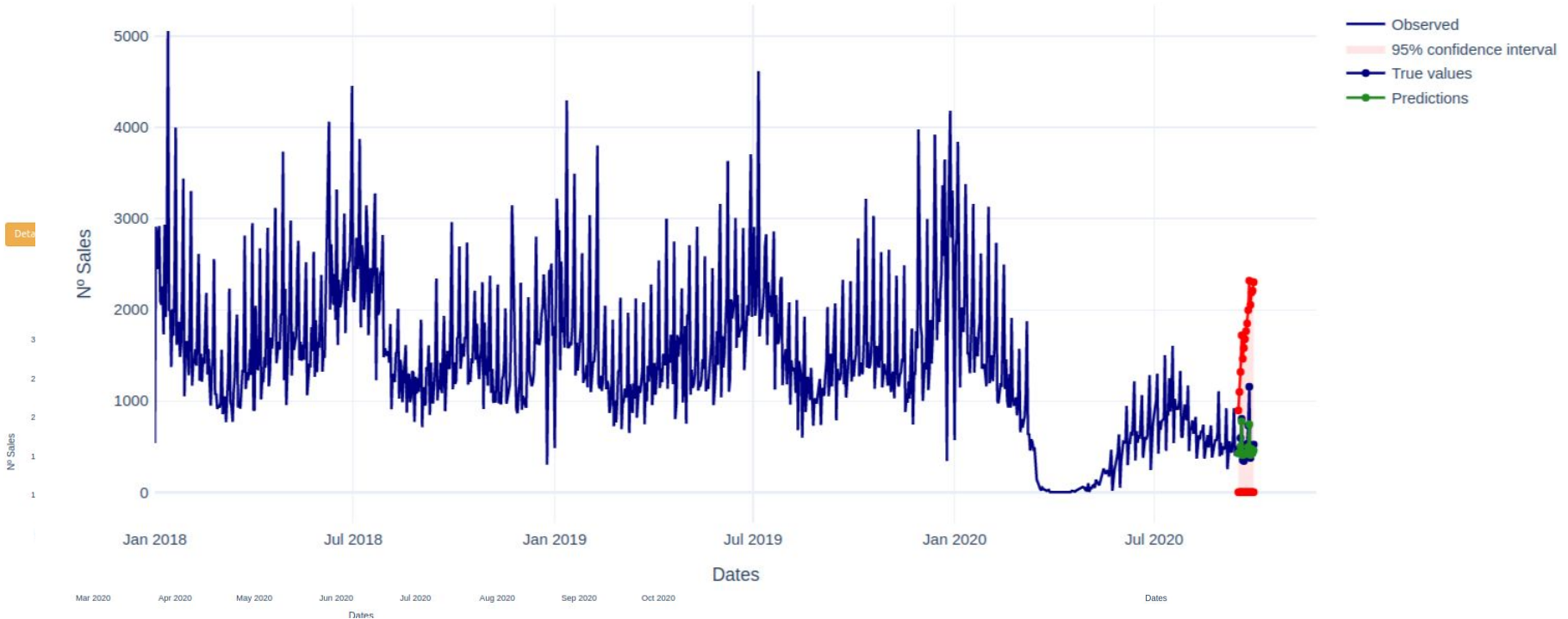
		Statistical criteria						
		Hig	Confidence Medium	Low	Hig	Discrimination Medium	Low	
Information	Reports			MSE				
				RMSE				
				MAPE				
				MAE	MSE	MdAE	MSSE	
				MdAE	MAPE	RMSE	RMSSE	
				MdAPE	MAE	sMAPE	MASE	
				RMSPE	mMdAPE		MdASE	
				RmsDSPE				
				MSSE				
				RMSSE				
User Criteria	Make Comparisons		MRAE	MAPE	MIRAE			
			GMRAE	MdAPE	GMRAE	U Thiel		
			MdRAE	RMSEPE	MAPE	sMAPE		
			sMAPE	RmdSPE	MdAPE			
			common sense	sMAPE	MAPE	MAPE	sMAPE	
				MAE				
			some meaning	MdAE	MdAPE	MAE		
				RMSPE	MdAPE		MdAE	MdRAE
			Understanding		RmdSPE			
					MSE			
User Criteria	not intuitive			RMSE				
				MSSE			MSSE	
			U-Theil	MRAE	RMSSE	MSE	RMSE	RMSE
				GMRAE	MASE		U-Theil	MASE
				MdRAE	MdASE			MdASE
					MdASE			

Details

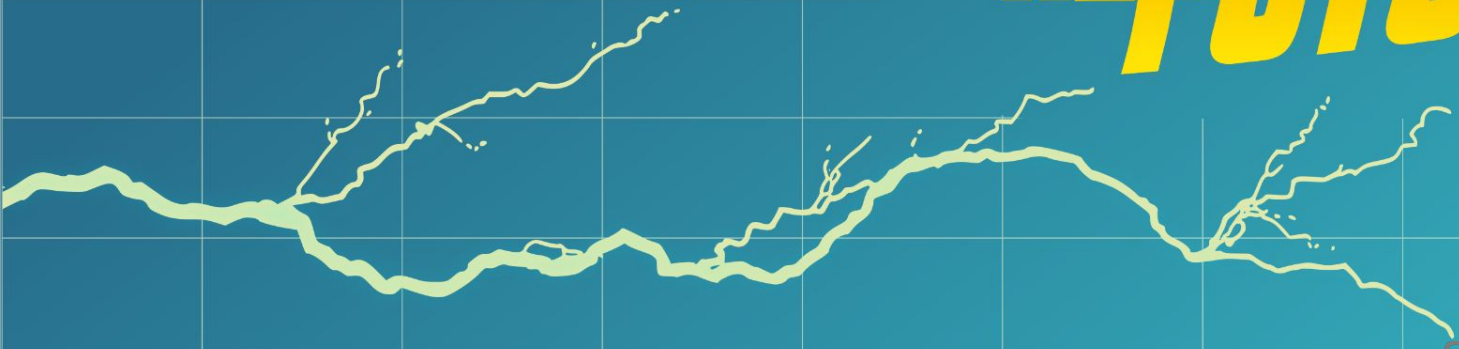
Json

Time Series Forecasting

Mean MAPE for predictions is 13.77% and MSPE coefficient of 11.77%



REGRESOR AL FUTURO



XGBOOST



REGRESOR AL FUTURO



d

[Gracias]

Do you want to know more?



info@apsl.net



[@apsl_web](https://twitter.com/apsl_web)