

Time Series Forecasting

The Marvik logo is centered on the page. It features the word "marvik" in a white, lowercase, sans-serif font. The letter "a" is stylized with a horizontal bar that extends to the left. The letter "i" has a dot that is a small blue square. The background consists of several overlapping, semi-transparent geometric shapes in various shades of blue, creating a modern, abstract pattern.

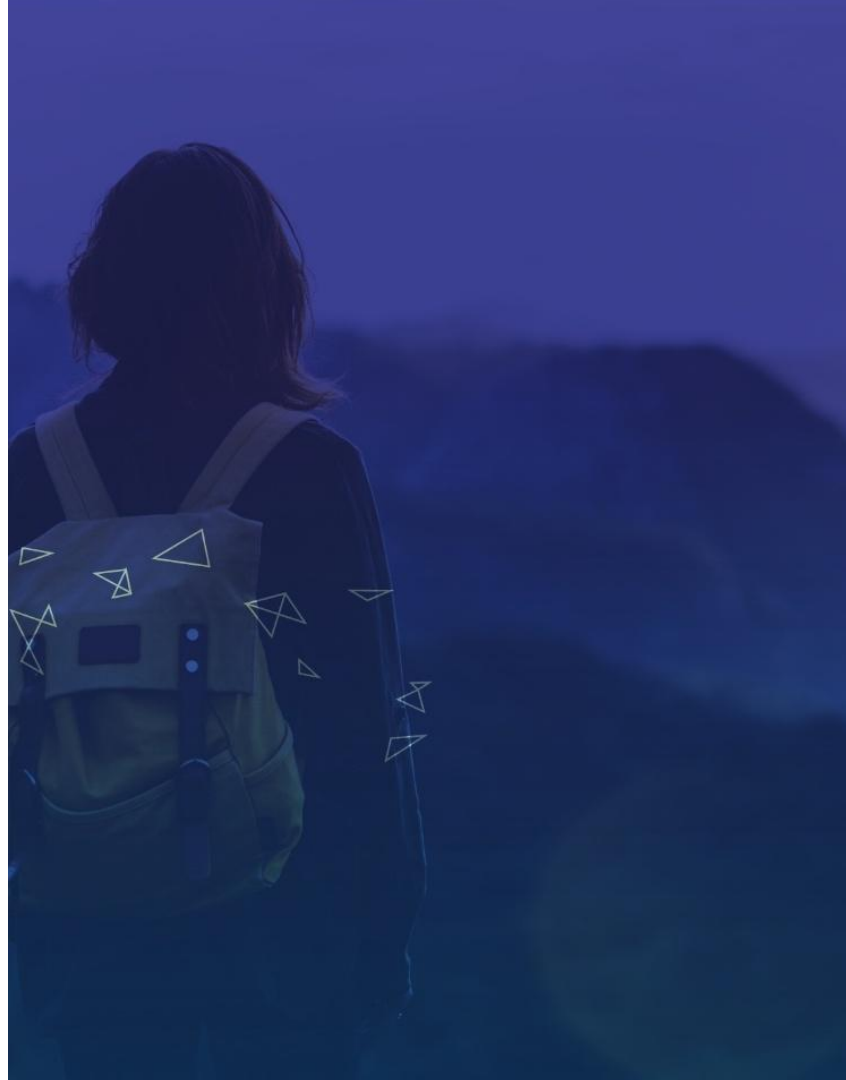
marvik.

AI solutions with business impact

marvik.ai

Agenda

- Motivación
- ¿Qué son las series temporales y qué significa su predicción?
- ¿Cómo abordar la problemática de predicción?
- Preparación de los datos
- Modelado de las series
- Estado del arte
- Resumen



Motivación

Motivación



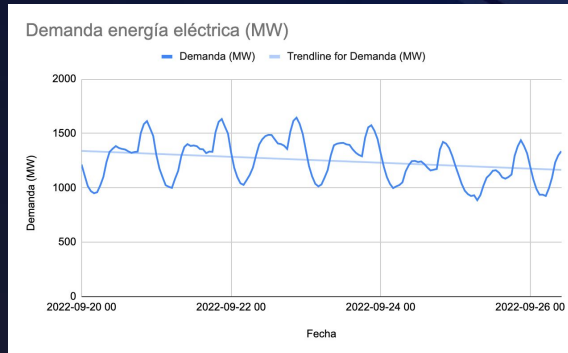
Ranking de problemas de los clientes de Google Cloud

**¿Qué son las series temporales y
qué significa su predicción?**

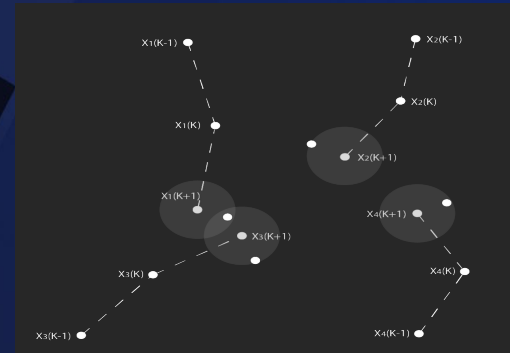
¿Qué son las series temporales y qué significa su predicción?

Series temporales

- Una serie temporal es un conjunto de datos que depende del tiempo. Es decir, que los datos se corresponden a un instante de tiempo.
- El concepto, se puede extender y puede aplicarse a casos donde la dependencia con el tiempo no es tan evidente (ej. Posición de un objeto en el espacio)
- Cuando hablamos de predicción, buscamos obtener un valor futuro en función de los valores históricos de la serie y otras características.



Demanda de energía eléctrica [ADME]



Tracking de múltiples objetos

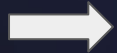
**¿Cómo abordar la problemática de
predicción?**

¿Cómo abordar la problemática de predicción?

- Análisis previo de los datos
 - Cuantificar la cantidad de datos disponibles
 - Analizar la complejidad de la serie
 - Evaluar todos los inputs posibles
 - Modelar el problema completo
- **Preparación de los datos**
 - Separar los datos de manera correcta
 - Identificar qué variables tomar en cuenta y cuáles no en los casos de series multivariadas
 - Creación de nuevas características
- **Elegir el/los modelo/s**
 - En base a los dos pasos anteriores, decidir cuáles modelos son prometedores
- Evaluar performance y re-entrenar
 - Una vez se siguieron los pasos anteriores, analizar la salida e identificar puntos de mejora y ajustar el modelo periódicamente

¿Cómo abordar la problemática de predicción?

- Análisis previo de los datos
 - Cuantificar la cantidad de datos disponibles
 - Analizar la complejidad de la serie
 - Evaluar todos los inputs posibles
 - Modelar el problema completo



- **Preparación de los datos**
 - Separar los datos de manera correcta
 - Identificar qué variables tomar en cuenta y cuáles no en los casos de series multivariadas
 - Creación de nuevas características



- **Elegir el/los modelo/s**
 - En base a los dos pasos anteriores, decidir cuáles modelos son prometedores
- Evaluar performance y re-entrenar
 - Una vez se siguieron los pasos anteriores, analizar la salida e identificar puntos de mejora y ajustar el modelo periódicamente

Preparación de los datos

Preparación de los datos

Vamos a ver:

- Tipos de datos aplicado al caso
- Cómo se suelen dividir los datos
- Alguna idea básica de implicancias que se tiene al hacer *feature engineering* dentro de series temporales
- Comparación con otros tipos de problemáticas
- Cosas a tener en cuenta cuando trabajamos con estos tipos de datos

Preparación de los datos

¿Qué tipos de características se tienen a la hora de trabajar con series temporales?

- Futuro conocido (eg. Invertido en publicidad)
- A partir del pasado (cualquier feature que se tenga al momento de predecir, que ya haya pasado eg. propia serie)
- Constantes (eg. sku): Sirven para el caso en que se predigan múltiples TS

Tiempo	SKU	Inversión en publicidad (US\$)	Ventas del mes anterior	Ventas	
2022-03-01	123456789	12000	None	40,000	
2022-04-01	123456789	12000	40,000	47,225	
2022-05-01	123456789	10000	47,225	54,449	
2022-06-01	123456789	10000	54,449	60,474	
2022-07-01	123456789	10000	60,474	66,498	
2022-08-01	123456789	12000	66,498	72,522	
2022-09-01	123456789	14000	72,522	?	} ¡Futuro!
2022-10-01	123456789	14000	?	?	

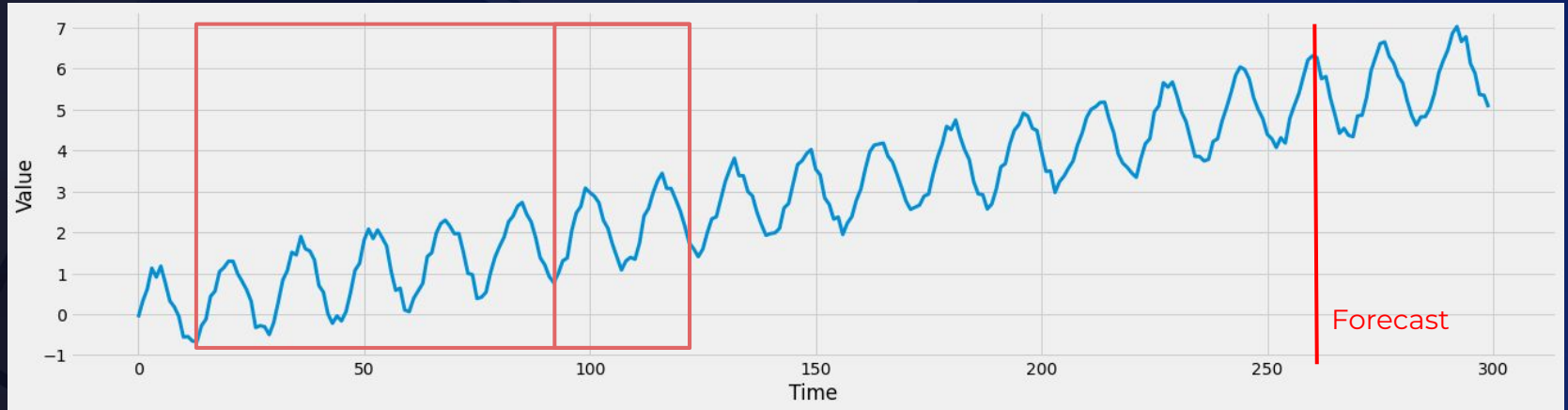
Constantes

Muestras pasadas

Futuro conocido

Preparación de los datos

- División de los datos a la hora de entrenamiento y testeo
- Se usan ventanas temporales
- Tener el cuidado de todo lo que se vaya computando, hacerlo en la ventana de tiempo correspondiente, **sin obtener información del futuro**



Preparación de los datos

- Se debe tener especial cuidado en evitar la fuga de información (*data leakage*)



Preparación de los datos

- Ejemplo *data leakage* haciendo ingeniería de datos

	Tiempo	Inversión en publicidad (U\$S)	Ventas del mes anterior	Promedio de ventas anteriores	Ventas (objetivo)
Train	2022-03-01	12000	None	56,861	40,000
	2022-04-01	12000	40,000	56,861	47,225
	2022-05-01	10000	47,225	56,861	54,449
	2022-06-01	10000	54,449	56,861	60,474
	2022-07-01	10000	60,474	56,861	66,498
Test	2022-08-01	12000	66,498	56,861	72,522
	2022-09-01	14000	72,522	63,636	79,747
	2022-10-01	14000	79,747	63,636	88,171

Preparación de los datos

- Posible forma de abordarlo

	Tiempo	Inversión en publicidad (U\$S)	Ventas del mes anterior	Promedio acumulado Ventas	Ventas (objetivo)
Train	2022-03-01	12000	None	-	40,000
	2022-04-01	12000	40,000	40,000	47,225
	2022-05-01	10000	47,225	-	54,449
	2022-06-01	10000	54,449	-	60,474
	2022-07-01	10000	60,474	-	66,498
Test	2022-08-01	12000	66,498	-	72,522
	2022-09-01	14000	72,522	-	?
	2022-10-01	14000	?	-	?

Preparación de los datos

- Posible forma de abordarlo

	Tiempo	Inversión en publicidad (U\$S)	Ventas del mes anterior	Promedio acumulado Ventas	Ventas (objetivo)
Train	2022-03-01	12000	None	-	40,000
	2022-04-01	12000	40,000	40,000	47,225
	2022-05-01	10000	47,225	43,612	54,449
	2022-06-01	10000	54,449	-	60,474
	2022-07-01	10000	60,474	-	66,498
Test	2022-08-01	12000	66,498	-	72,522
	2022-09-01	14000	72,522	-	?
	2022-10-01	14000	?	-	?

Preparación de los datos

- Posible forma de abordarlo

	Tiempo	Inversión en publicidad (U\$S)	Ventas del mes anterior	Promedio acumulado Ventas	Ventas (objetivo)
Train	2022-03-01	12000	None	-	40,000
	2022-04-01	12000	40,000	40,000	47,225
	2022-05-01	10000	47,225	43,612	54,449
	2022-06-01	10000	54,449	47,225	60,474
	2022-07-01	10000	60,474	50,537	66,498
	2022-08-01	12000	66,498	53,729	72,522
Test	2022-09-01	14000	72,522	-	?
	2022-10-01	14000	?	-	?

Preparación de los datos

Comparación entre caso de clasificación vs series temporales

	ML on tabular data (regression & classification)	ML on tabular data (forecasting)
Train/test split	Random allocation.	Split by time.
Creating the feature and target	Can pre-compute features and target before predict time.	Features built from target created "on demand" at predict time for test set.
Prediction	Only the trained model required at predict time.	Need trained model & training set at predict time.
Feature engineering		Time series specific feature engineering and data leakage issues.

Preparación de los datos

Resumen

- Analizar los datos previamente. Cuanto más lógica de negocio se tenga mejor!
- Como se mencionó, el leakage es un riesgo cuando se crean características a partir de la serie objetivo, así como de otras variables que también se tiene la información a futuro
- Solo usar los datos que se tienen hasta el momento que se quiere predecir

Modelado

Modelado

Si bien existen varias formas de predecir una serie temporal, dependiendo de la naturaleza de la serie, la información a priori que se tenga y los requerimientos, va a determinar qué tipo de modelo se utilice.

Métodos clásicos estadísticos

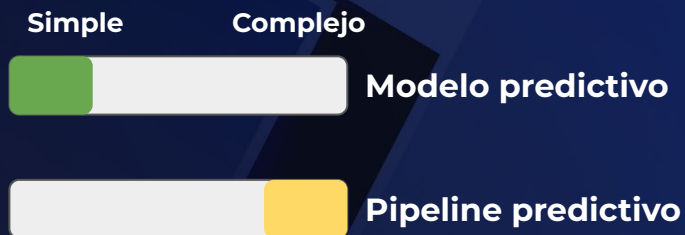
- Se modela la señal con distribuciones conocidas y parámetros a estimar
 - MAP, BLUE, MLE, etc
- Modelos
 - ETS, **ARIMA**, **PROPHET**, Kalman Filters

Machine Learning y Deep Learning

- Árboles de decisión
- **Redes convolucionales**
- **Redes Recurrentes**
- **Modelos de atención**

Métodos clásicos

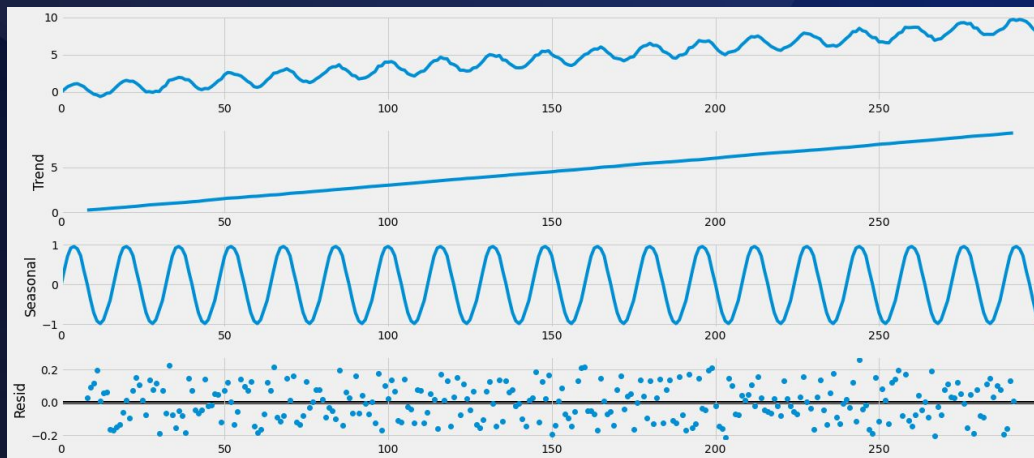
- La principal suposición del modelado clásico es que las series son estacionarias
- Estacionariedad implica:
 - Las estadísticas de primer y segundo orden son constantes
- Para poder aplicar los modelos, suele ser necesario un preprocesamiento y adaptación de los datos complejo en función de la complejidad de las series



ARIMA

"Auto Regressive Integrated Moving Average"

- descompone la serie en tres componentes
- predice basándose en muestras pasadas



Un modelo $ARIMA(p,d,q)$ se descompone de la siguiente manera:

- **AR(p):** p muestras pasadas para modelar el proceso en tiempo actual
- **I(d):** d la mínima cantidad de diferencias para hacer la serie estacionaria
- **MA(q):** El proceso se supone producido por una combinación de q muestras observadas de ruido blanco

PROPHET

Modelo hecho por Facebook y open source

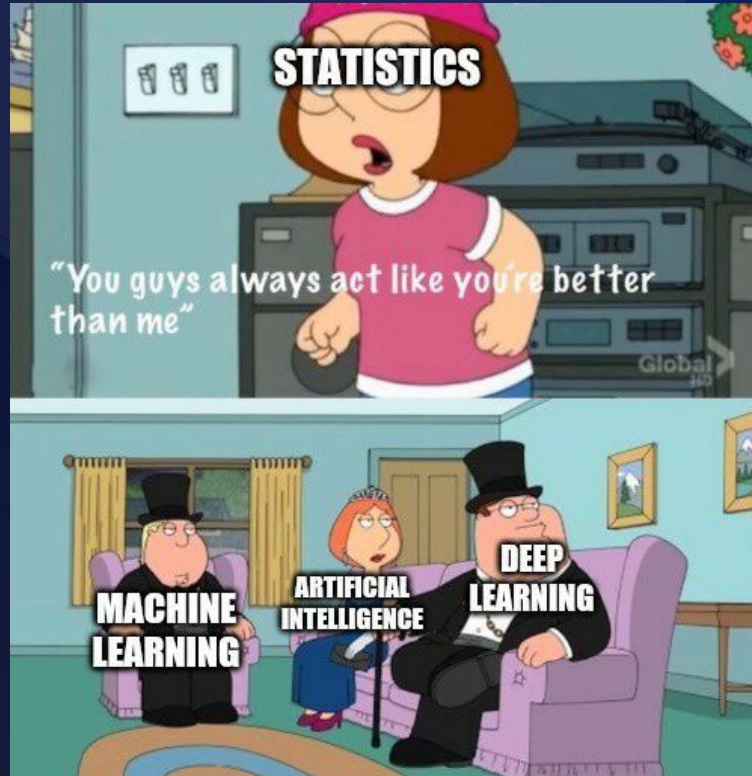
Prophet modela la serie tomando las siguientes componentes

- **Tendencia:** Modela los cambios no periódicos
- **Estacionalidad:** Modela los cambios periódicos
- **Componentes de recesos:** Contribuye información acerca de eventos y vacaciones

Características

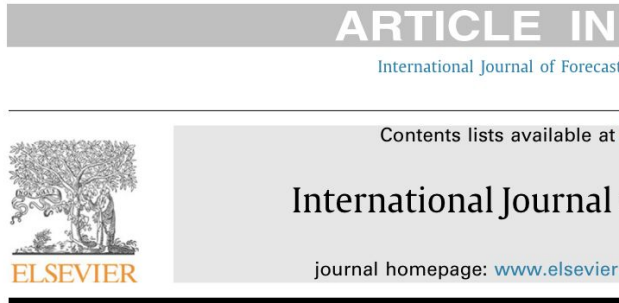
- Demuestra ser mejor método que ARIMA en casos donde se ven efectos humanos como vacaciones, fechas y otros efectos
- Es más robusto a datos faltantes
- Es un framework completo, no solo un modelo

Machine learning para predicción



Machine learning para predicción

- Las M1, M2, ..., M6 competencias



M5 accuracy competition: Results, findings, and conclusions

Spyros Makridakis^b, Evangelos Spiliotis^{a,*}, Vassilios Assimakopoulos^a

^a Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, Greece

^b Institute For the Future, University of Nicosia, Cyprus

ARTICLE INFO

ABSTRACT

4.3. Key findings

The main findings related to the performance of the top five methods are summarized as follows.

Finding 1: Superior performance of ML methods. Over many years, empirical studies have demonstrated that simple methods are as accurate as complex or statistically sophisticated methods (Makridakis et al., 2020c). Limited data availability, inefficient algorithms, the need for preprocessing, and restricted computational power are just some of the factors that reduce the accuracy of ML

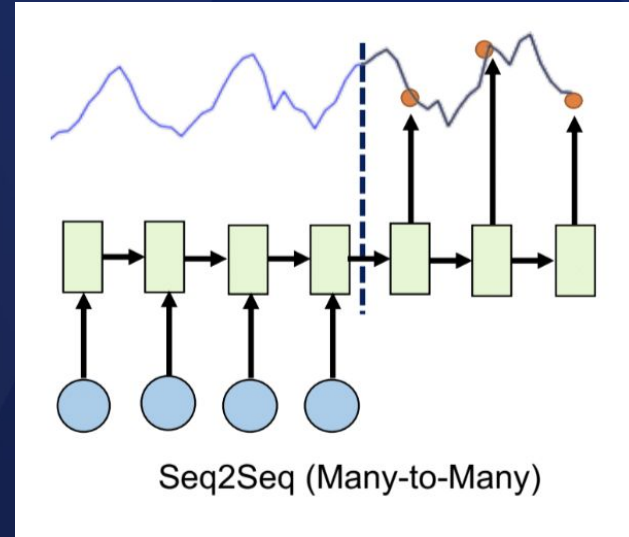
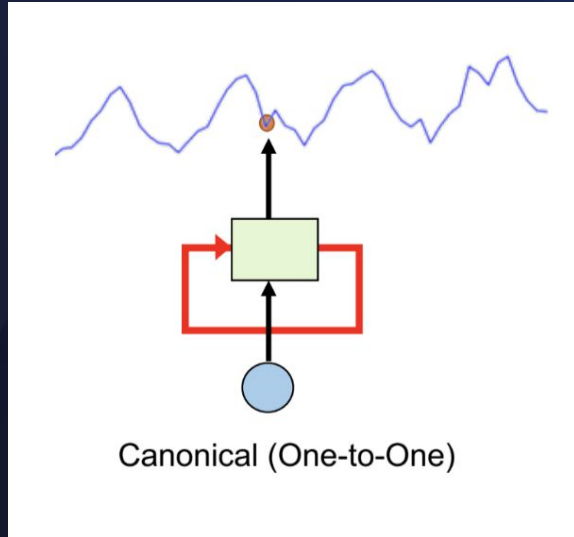
[Source](#)

Machine learning para predicción

- Cada vez más utilizadas
- Varias opciones
- Suelen ser robustos frente a distintos tipos de series temporales
- Algunas versiones simples de los distintos modelos suelen ser análogos a métodos clásicos
- Suelen aprender patrones que en los métodos clásicos es necesario resolver con información a priori

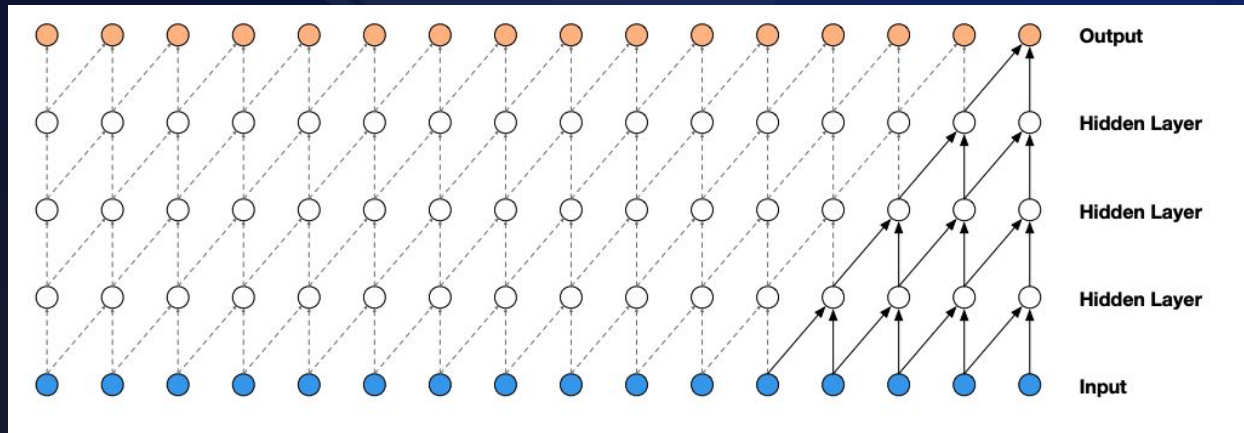


Machine learning para predicción



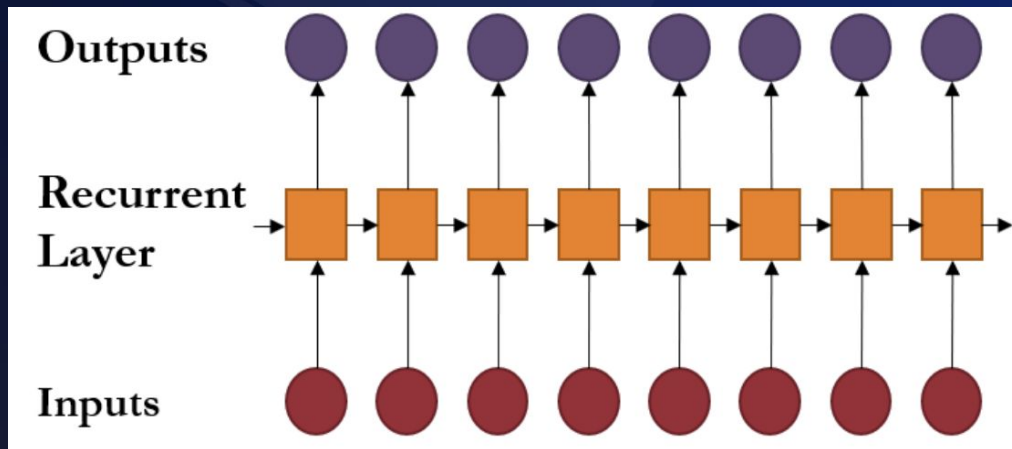
Temporal Convolutional Neural Networks

- Se suele usar *Causal Convolution* para evitar leakage
- Suelen ser más rápidos de entrenar en comparación con las redes recurrentes
- Asumen que las relaciones son temporalmente invariantes
- Requieren muchas capas o filtros grandes para aumentar el "campo receptivo"



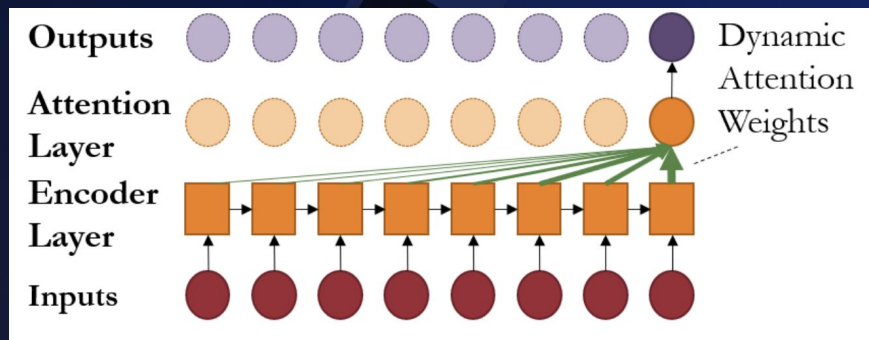
RNN

- Históricamente usadas para modelado secuencial
- Internamente cuentan con un estado que incluye información del pasado resumida (memoria)
- Comparables a filtros bayesianos dado su mantenimiento de memoria
- Performance superior frente a las CNNs



MODELOS DE ATENCIÓN

- El mecanismo de atención también llevó a mejoras considerables en la dependencia de largo término
- Agregan características temporales usando pesos dinámicamente creados
 - Permite que la red se enfoque en pasos del pasado que sean significantes
- Dos beneficios clave:
 - Son capaces de dirigir la atención directamente a cualquier evento significativo que pase (ej: feriados)
 - Pueden aprender dinámicas temporales en régimen



Comparación general

Comparación

- Pipeline complejo + Modelo simple
 - Suele ser el caso de los modelos clásicos
 - Pros
 - Son transparentes
 - El modelo suele ser interpretable y sencillo
 - Se puede interferir en las predicciones
 - Cons
 - Se suele necesitar un conocimiento profundo de negocio
 - El pipeline puede llegar a ser realmente complejo
 -
- Modelo complejo + Pipeline simple
 - Suele ser el caso de redes neuronales
 - Pros
 - Suelen ser robustos
 - Tienen mejor performance en casos más complejos
 - Logran encontrar patrones en caso realmente complejos
 - Cons
 - Suelen ser poco interpretables
 - No se tiene control sobre las predicciones

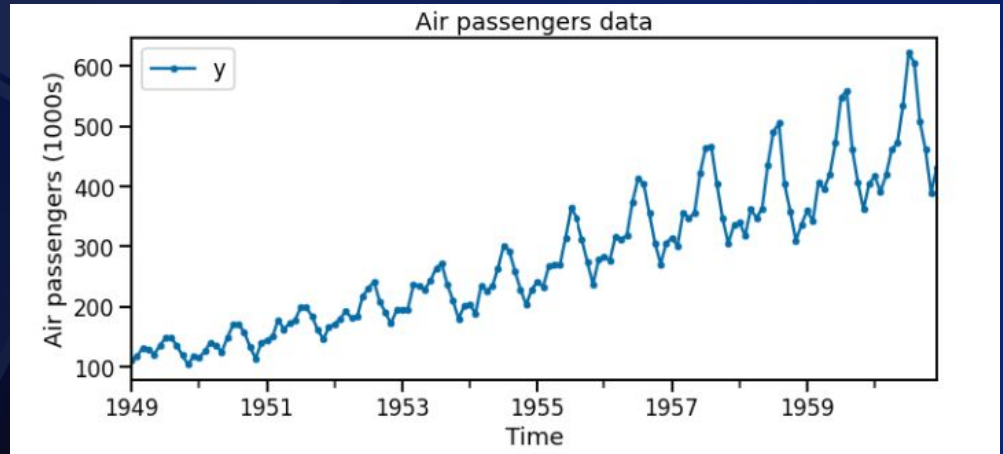
¿Qué modelo usar?



Cuándo usar enfoque clásico

Características de la serie

- Fuerte estacionalidad (seasonality)
- Poca cantidad de series
- Series no correlacionadas
- Sin esparsidad o intermitencia
- Pocas o sin variables exógenas



Cuándo usar enfoque de ML

Características de la serie

- Gran cantidad de series correlacionadas
- Gran esparsidad e intermitencia
- Variables exógenas (precios, promos, etc)
- Muchos patrones de estacionalidad
- Estructura jerárquica

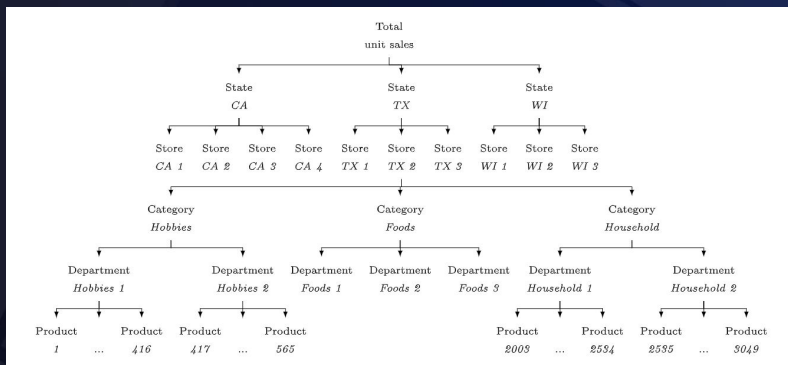


Fig. 1. Grouped time series used in the M5 competition. The data can be aggregated in 12 different levels using either location (state and store) or product-related information (category and department).

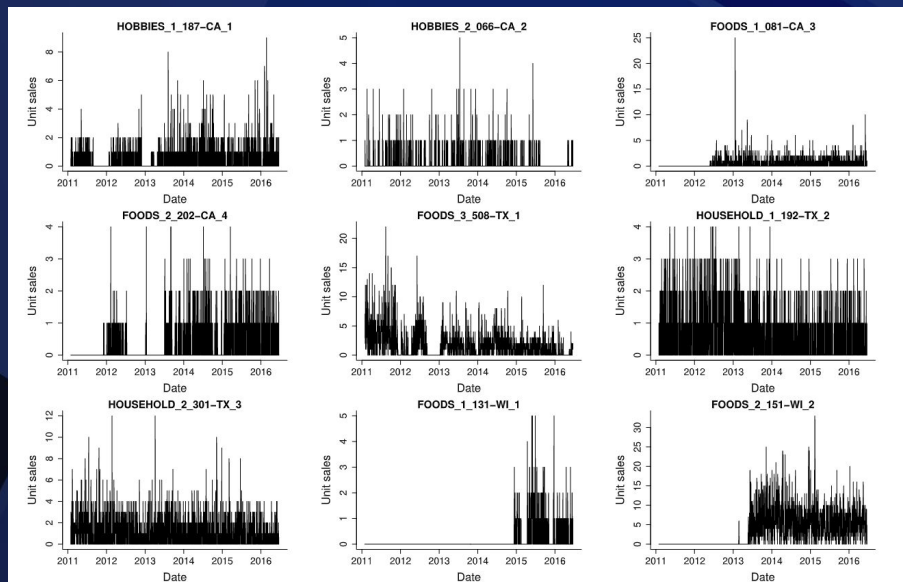


Fig. 4. A sample of series reporting the daily unit sales recorded at the product-store level. Series "HOBBIES_1_187-CA_1" refers to the sales recorded for the 187th product of the *Hobbies 1* department sold at the first store in CA, "HOBBIES_2_066-CA_2" refers to the sales recorded for the 66th product of the *Hobbies 2* department sold at the second store in CA, and so on.

Estado del arte

Estado del arte

- Definir el estado del arte dentro de time series forecasting no es trivial
- Las decisiones a tomar siguen siendo **muy dependientes del problema**
- Es necesario analizar los datos y objetivos buscados para seleccionar el correcto pipeline y modelo
- Para elegir el estado del arte en time series forecasting, se siguieron los siguientes criterios
 - Modelo interpretable
 - Código disponible
 - Aplicable en un amplio espectro de datos

Estado del arte

International Journal of Forecasting 37 (2021) 1748–1764



Contents lists available at [ScienceDirect](#)

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast



Temporal Fusion Transformers for interpretable multi-horizon time series forecasting

Bryan Lim^{a,*}, Sercan Ö. Arık^b, Nicolas Loeff^b, Tomas Pfister^b

^a University of Oxford, UK

^b Google Cloud AI, USA



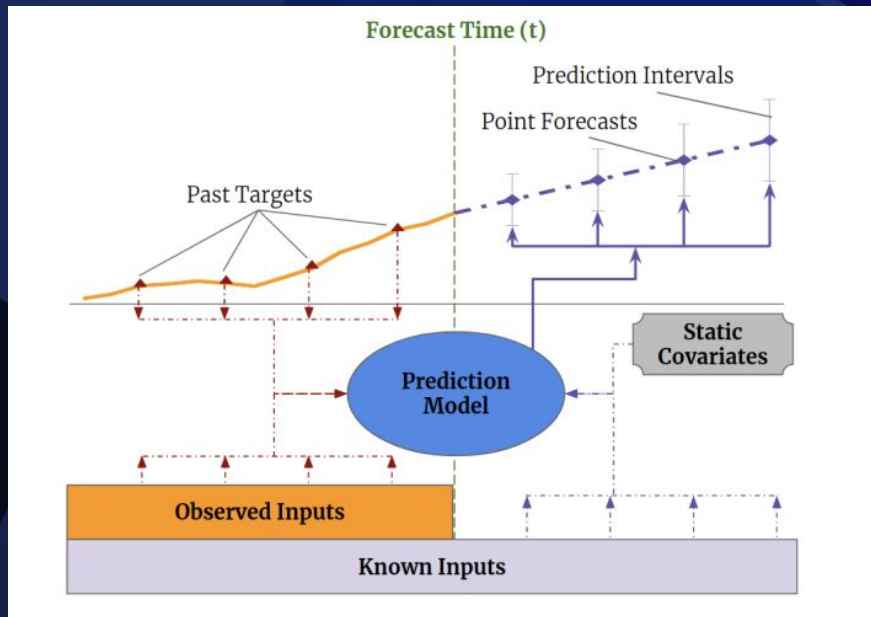
Datos

- Se diseñó para que el modelo pudiera trabajar con una amplia diversidad de datos
- Se diseñó para utilizar:
 - Múltiples tipos temporales de entradas
 - Múltiples tipo de entradas (numéricas, categóricas, no estructurada)
 - Muchas entradas que terminan siendo ruido
 - **Muchas** series de tiempo involucradas

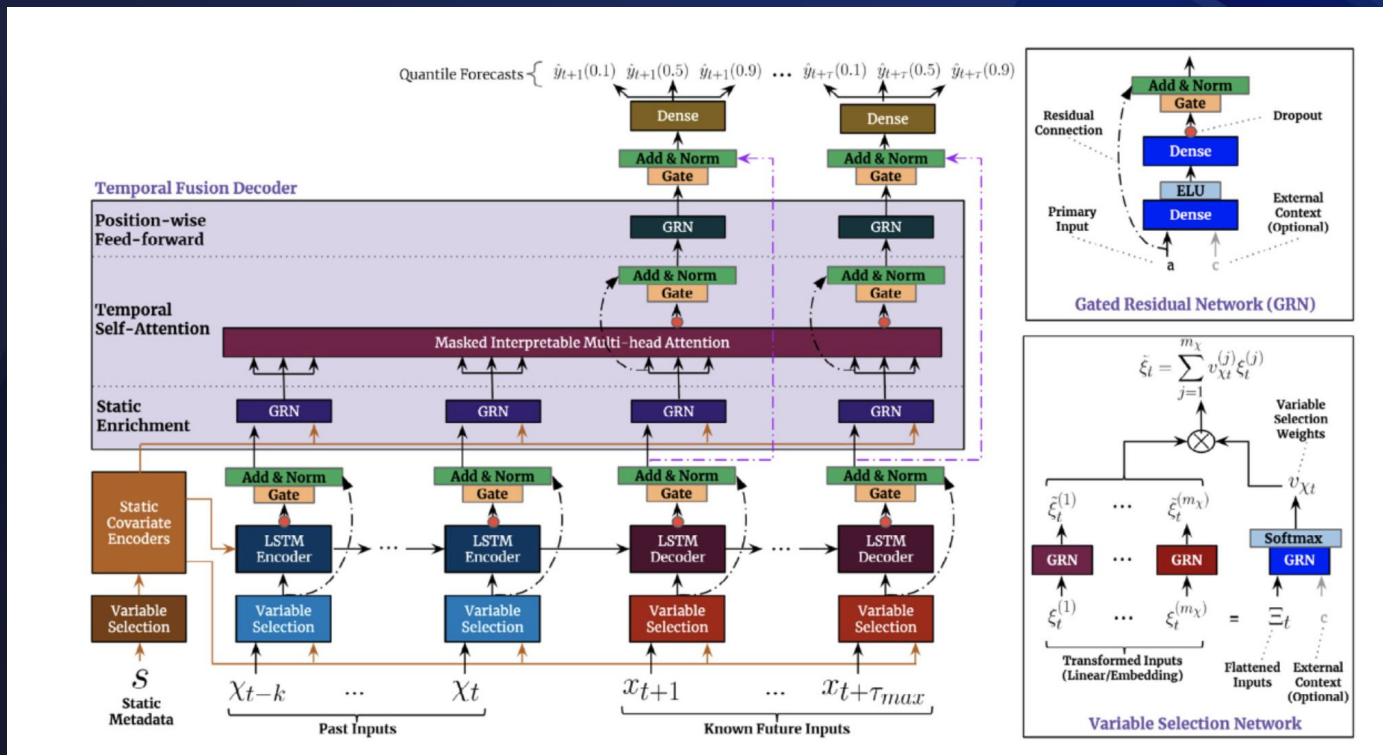


Características buscadas

- Modelo interpretable para confiabilidad de los humanos
- Modelo capaz de "apagar" entradas que fuesen ruido (o no relevantes)
- Modelo capaz de generalizar y escalar
- Resiliencia frente a datos faltantes
- Manejo de distinta longitud de datos
- **Entrenamiento completo**

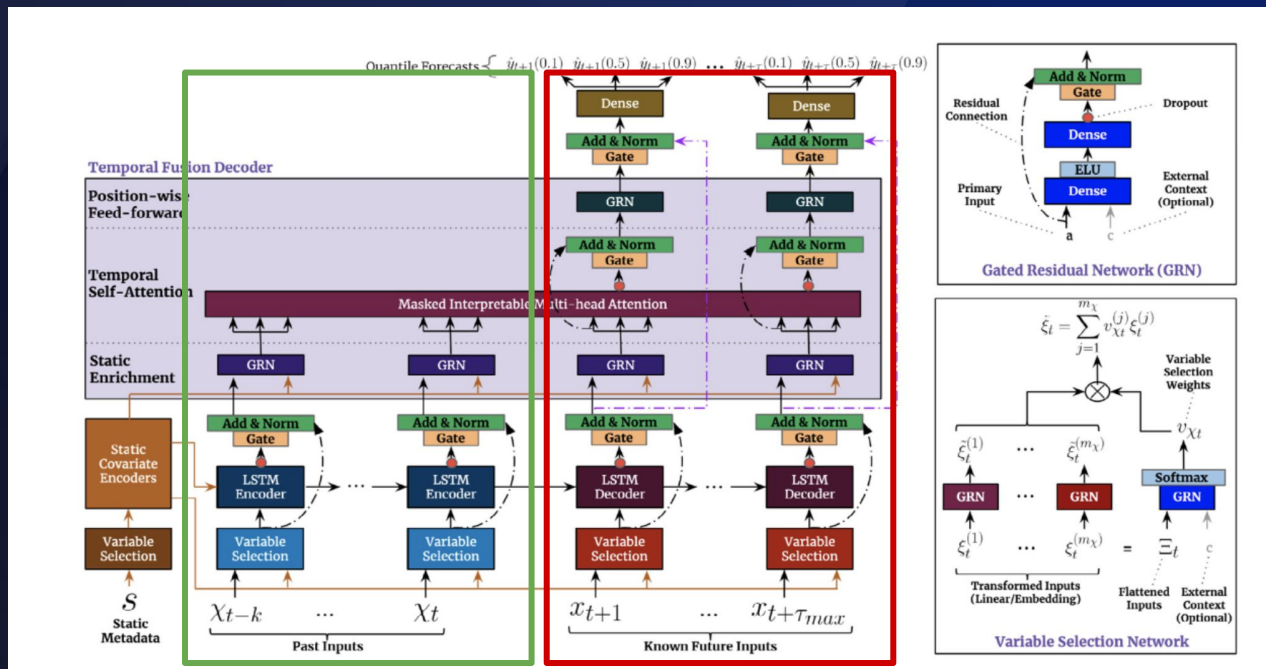


Modelo



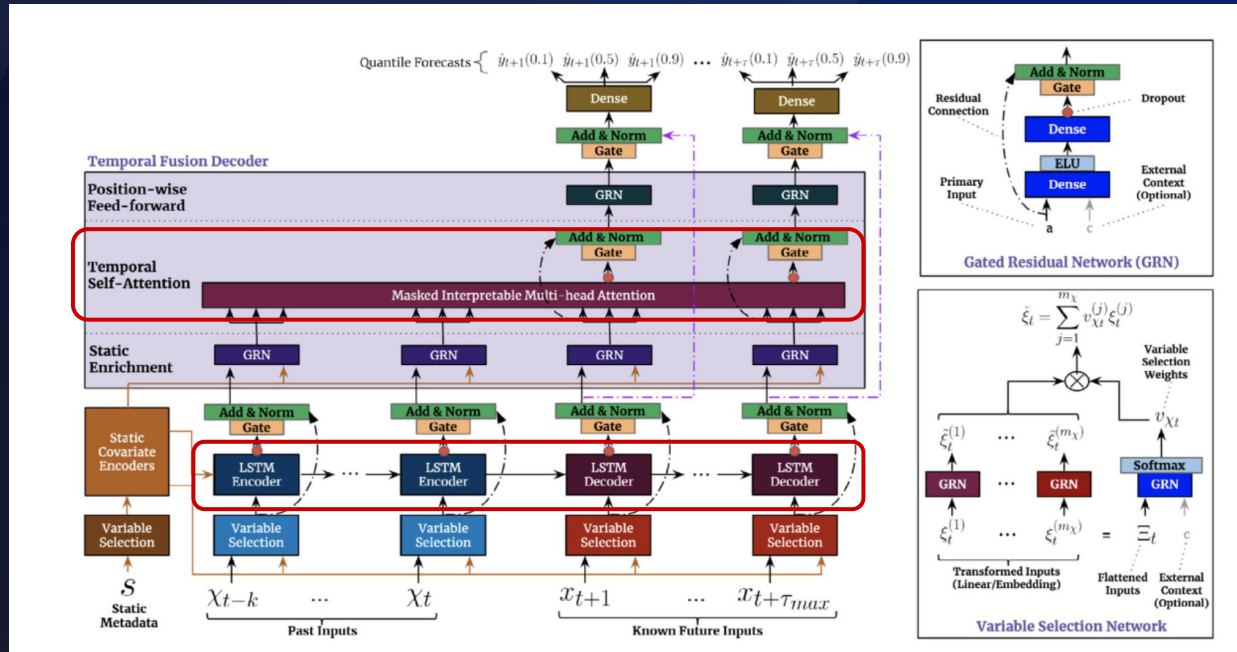
Modelo

- Modelo seq2seq



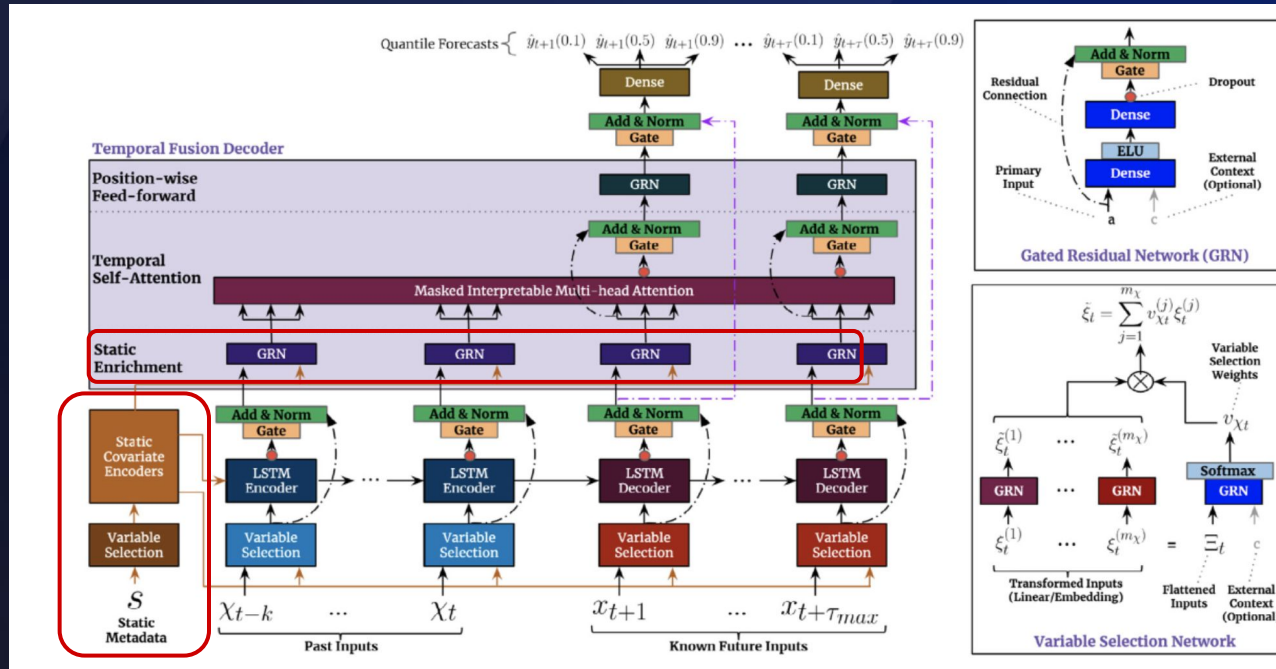
Modelo - Procesamiento temporal

- Capas de atención: Para modelar pasado distante
- Capas recurrentes: Para modelar dinámica local

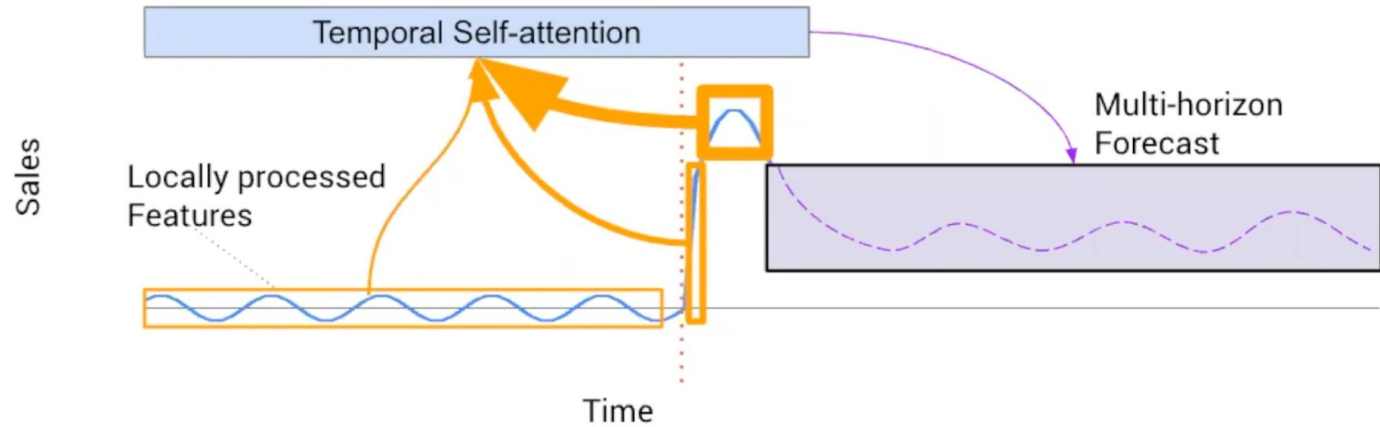


Modelo - Entradas estáticas

- Codificación de entradas estáticas
 - Utiliza vectores de contexto para condicionar las dinámicas temporales

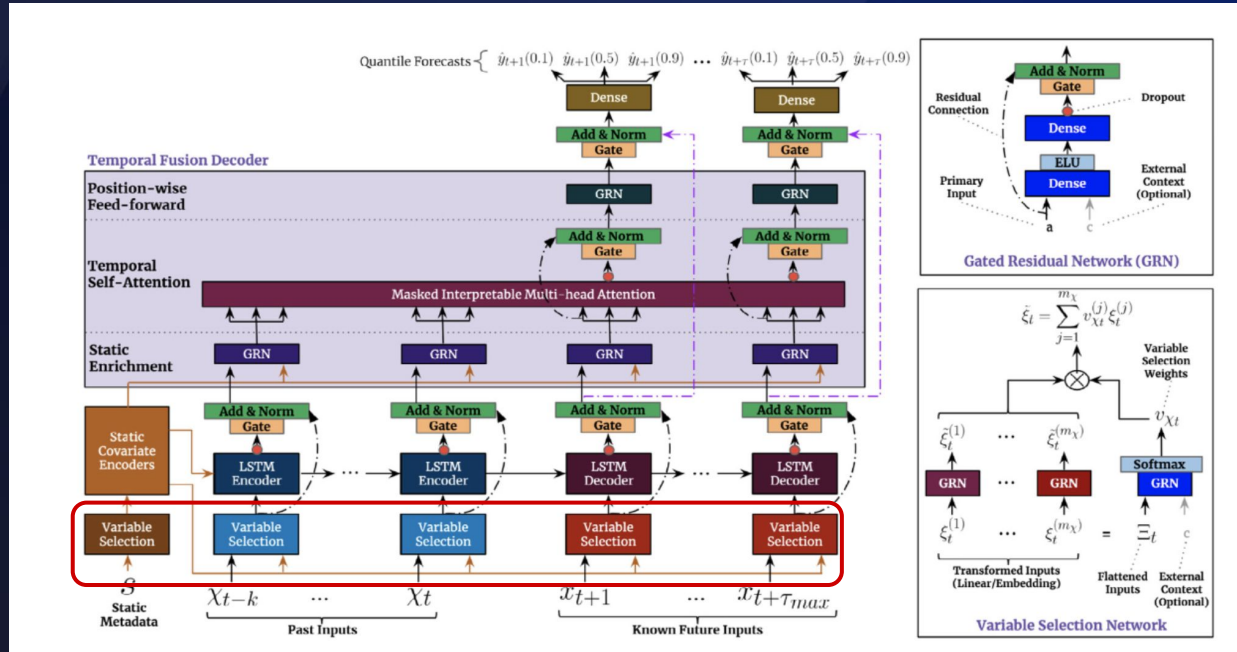


Modelo



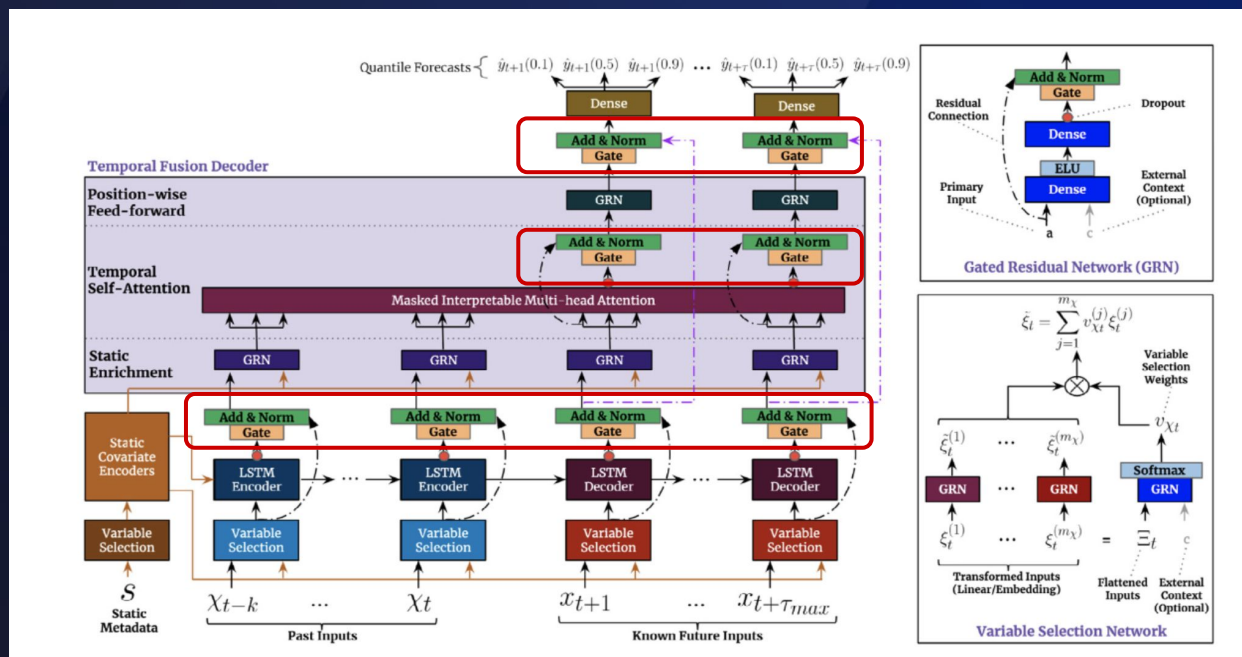
Modelo - Selección de variables

- Selección de variables importantes en cada instante de tiempo
- Usan el vector de contexto para "aumentar/disminuir predicciones"



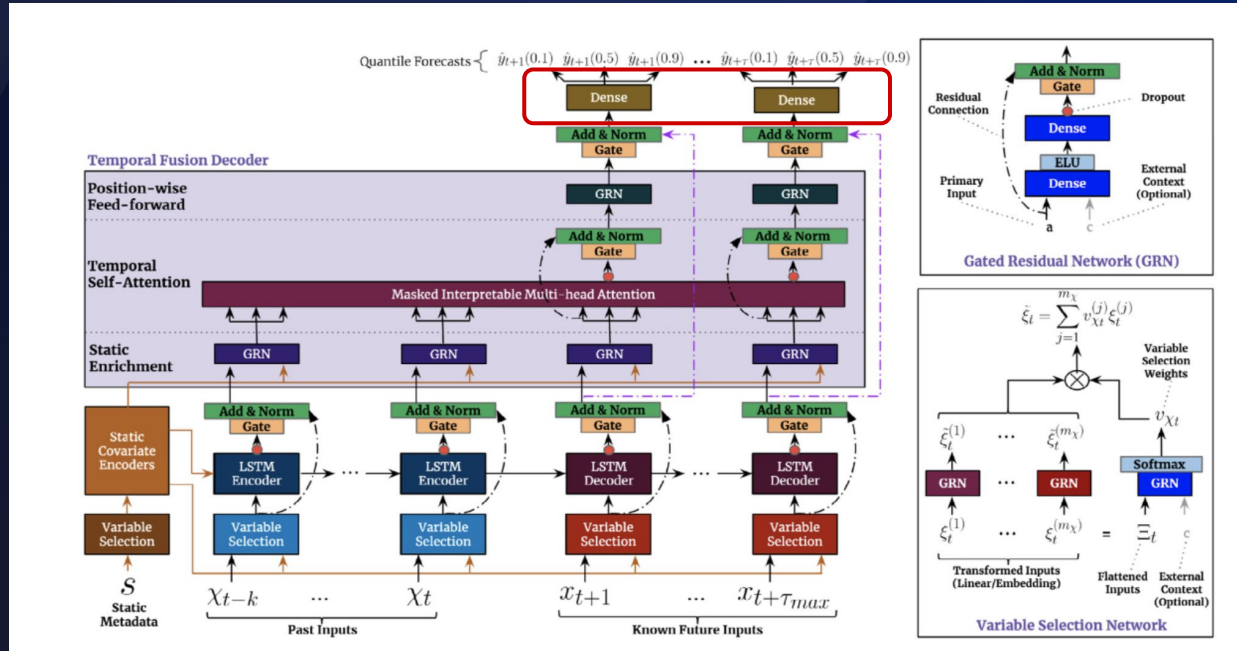
Modelo - Variación de complejidad

- Se agregaron *gates* para adaptar la complejidad del modelo!



Modelo - Predicción

- Predicción de intervalos vía predicción de cuantiles



Modelo

- Lograron crear un modelo que adapta su complejidad de manera automática, dependiendo del tipo de serie y sus entradas
- SOTA en datasets variados
- El uso de los componentes explicados, logra que el modelo sea interpretable
 - Detección de variables globalmente importantes para el problema
 - Patrones temporales persistentes
 - Eventos significativos

	ARIMA	ETS	TRMF	DeepAR	DSSM
Electricity	0.154 (+180%)	0.102 (+85%)	0.084 (+53%)	0.075 (+36%)	0.083 (+51%)
Traffic	0.223 (+135%)	0.236 (+148%)	0.186 (+96%)	0.161 (+69%)	0.167 (+76%)
	ConvTrans	Seq2Seq	MQRNN	TFT	
Electricity	0.059 (+7%)	0.067 (+22%)	0.077 (+40%)	0.055*	
Traffic	0.122 (+28%)	0.105 (+11%)	0.117 (+23%)	0.095*	

(a) P50 losses on simpler univariate datasets.

	ARIMA	ETS	TRMF	DeepAR	DSSM
Electricity	0.102 (+278%)	0.077 (+185%)	-	0.040 (+48%)	0.056 (+107%)
Traffic	0.137 (+94%)	0.148 (+110%)	-	0.099 (+40%)	0.113 (+60%)
	ConvTrans	Seq2Seq	MQRNN	TFT	
Electricity	0.034 (+26%)	0.036 (+33%)	0.036 (+33%)	0.027*	
Traffic	0.081 (+15%)	0.075 (+6%)	0.082 (+16%)	0.070*	

(b) P90 losses on simpler univariate datasets.

	DeepAR	CovTrans	Seq2Seq	MQRNN	TFT
Vol.	0.050 (+28%)	0.047 (+20%)	0.042 (+7%)	0.042 (+7%)	0.039*
Retail	0.574 (+62%)	0.429 (+21%)	0.411 (+16%)	0.379 (+7%)	0.354*

(c) P50 losses on datasets with rich static or observed inputs.

	DeepAR	CovTrans	Seq2Seq	MQRNN	TFT
Vol.	0.024 (+21%)	0.024 (+22%)	0.021 (+8%)	0.021 (+9%)	0.020*
Retail	0.230 (+56%)	0.192 (+30%)	0.157 (+7%)	0.152 (+3%)	0.147*

(d) P90 losses on datasets with rich static or observed inputs.

Resumen

Resumen

- Se debe tener especial cuidado a la hora de preparar los datasets y crear características nuevas
- La elección del modelo depende de varios factores
- En general se ha comprobado que ponderar estimaciones de varios modelos suele tener un mejor desempeño que utilizar solo uno
- Las redes neuronales siguen evolucionando para abarcar las características positivas que se tienen con los modelos clásicos y seguir mejorando la predicción
- Trade off pipeline vs modelo
- Naturalmente, cuanto más grande sea el horizonte de predicción más incertidumbre se agrega a la predicción
- Tomar un modelo que parezca que tenga sentido y probar un par de variaciones del modelo

Thank you!

| We are ready to help you!

✉ hello@marvik.ai

🌐 marvik.ai

marvik.