

# MODELO PREDICTIVO SP500

Proyecto de predicción de Machine Learning y Deep Learning del S&P500,

Adrián Ruiz Donado

# CONTENIDO

**01**

Resumen

**02**

Objetivos

**03**

EDA

**04**

Prophet y Arima

**05**

LSTM

**06**

Modelos  
Multivariantes

**07**

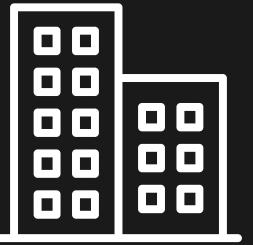
Conclusión

**08**

Contacto



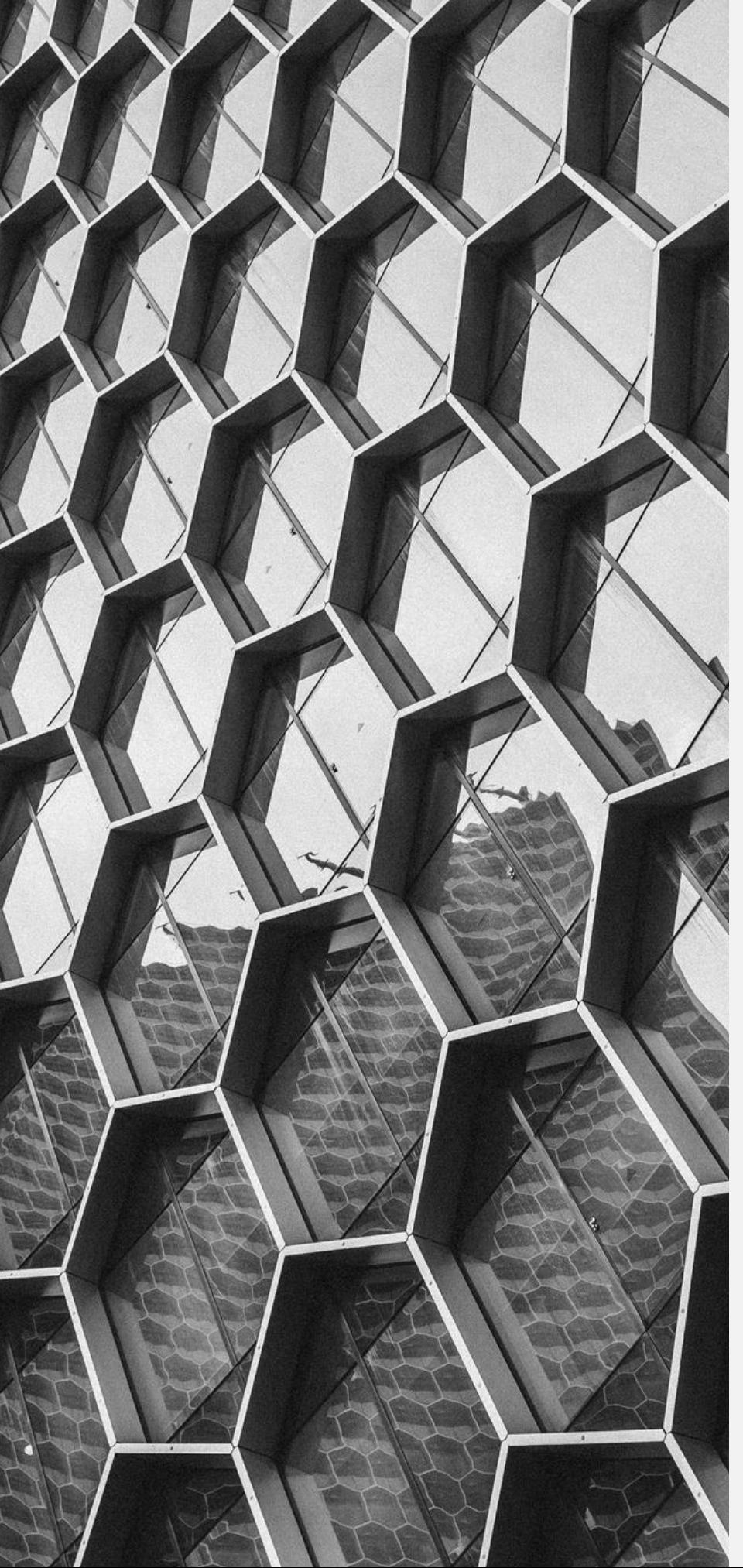
# RESUMEN DEL PROYECTO



Desde hace décadas predecir los índices, valores y divisas ha sido la obsesión de muchos inversores



Desarrollar modelos que simplemente reduzcan el error o ayuden en la toma de decisiones tiene un valor incalculable



# OBJETIVOS



## Objetivo n° 1

Predecir el S&P500



## Objetivo n° 2

Errores iguales o cercanos al media de variación diaria



## Objetivo n° 3

Optimizar lo máximo posible el modelo

# HIGHLIGHTS

## Datasets

- o 1 dataset S&p500
- o Librería Yahoo Finance

## Estado datos

- o Alta cardinalidad
- o Time Series
- o Columnas superfluas

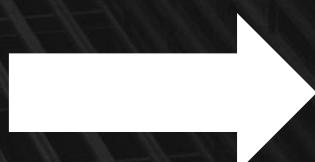
## Limpieza y tratamiento de datos

- o Adaptación a Prophet
- o Fechas a Datetime
- o Escalados según modelo

01



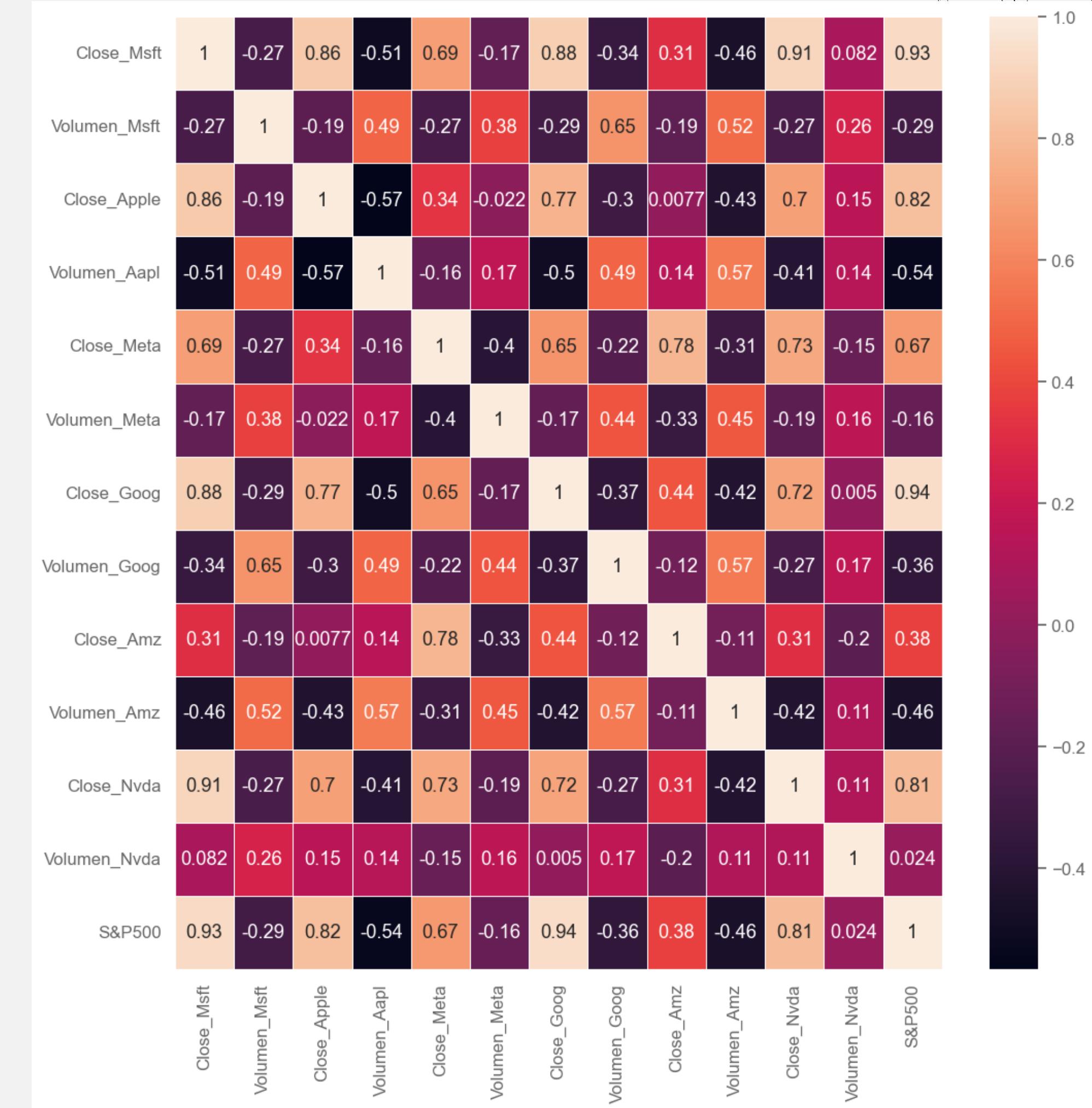
02



03

# EDA

- S&P500 tiene una fuerte relación con las FANG (22% de su capitalización)
- Los volúmenes de negociación tienen una correlación negativa notable con los índices.
- Se han elegido las features más representativas de manera cualitativa

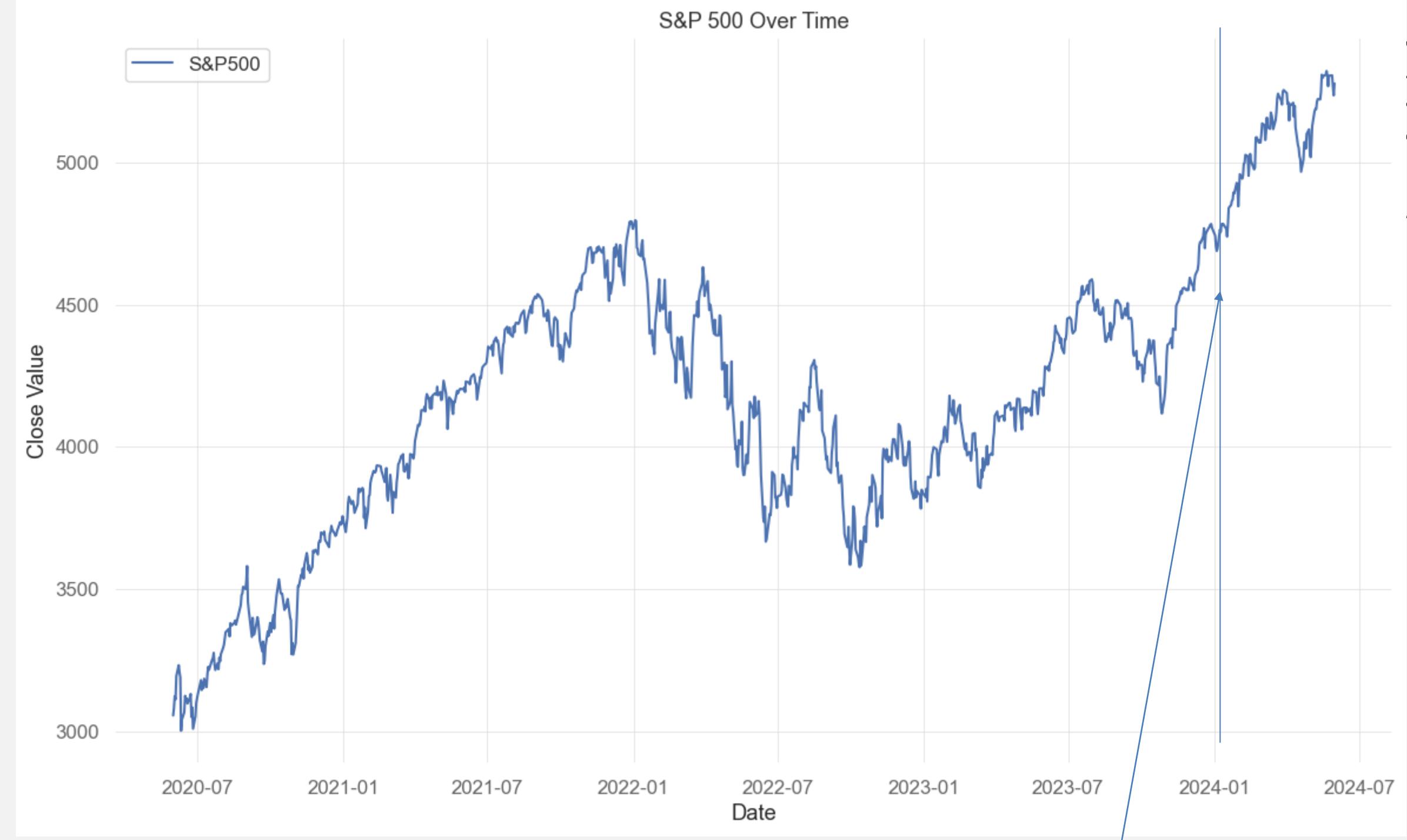


# EL ÍNDICE

- Tendencia alcista
- Periodo de 4 años



- No es estacionaria
- No se ve seasonality



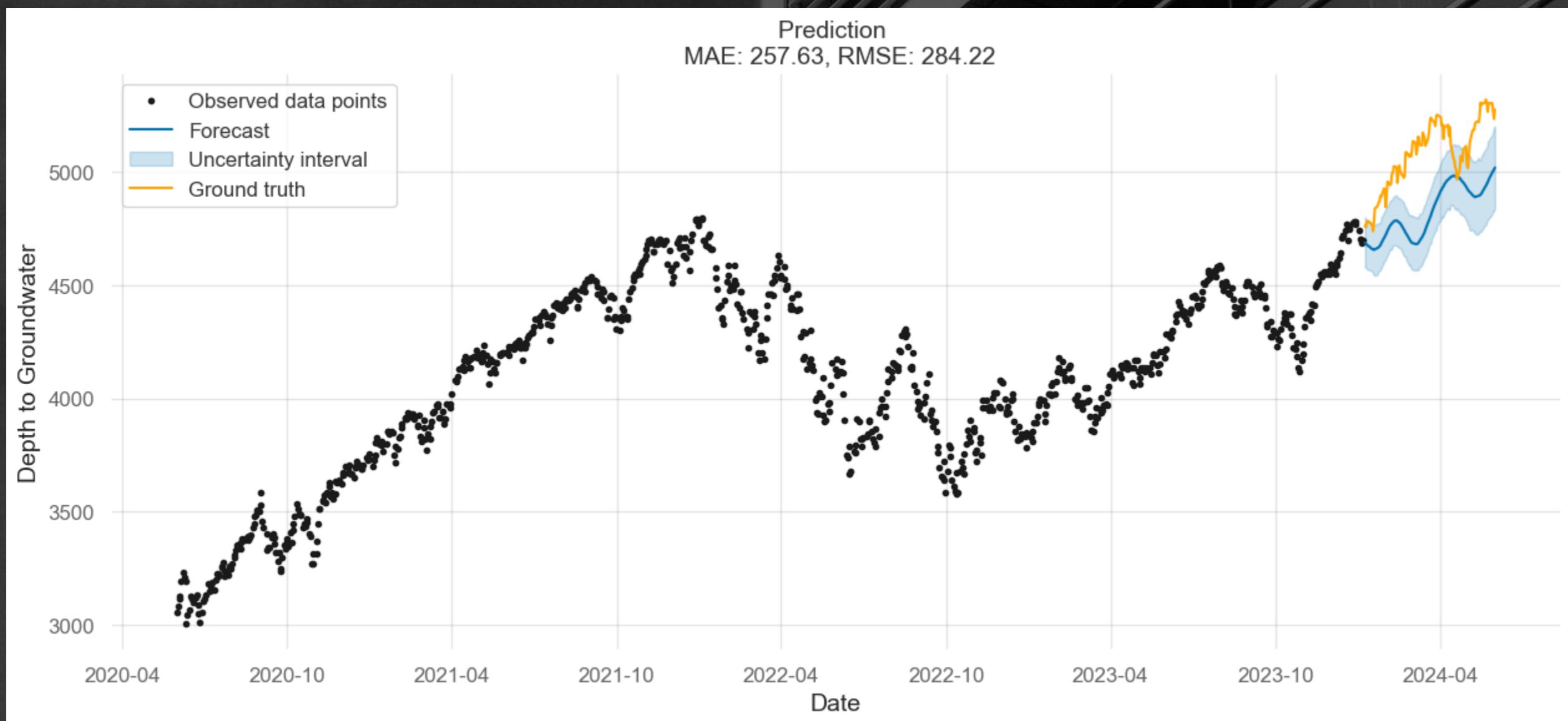
Predicciones corto plazo

90% train y 10 % test

# MODELOS UNIVARIANTES

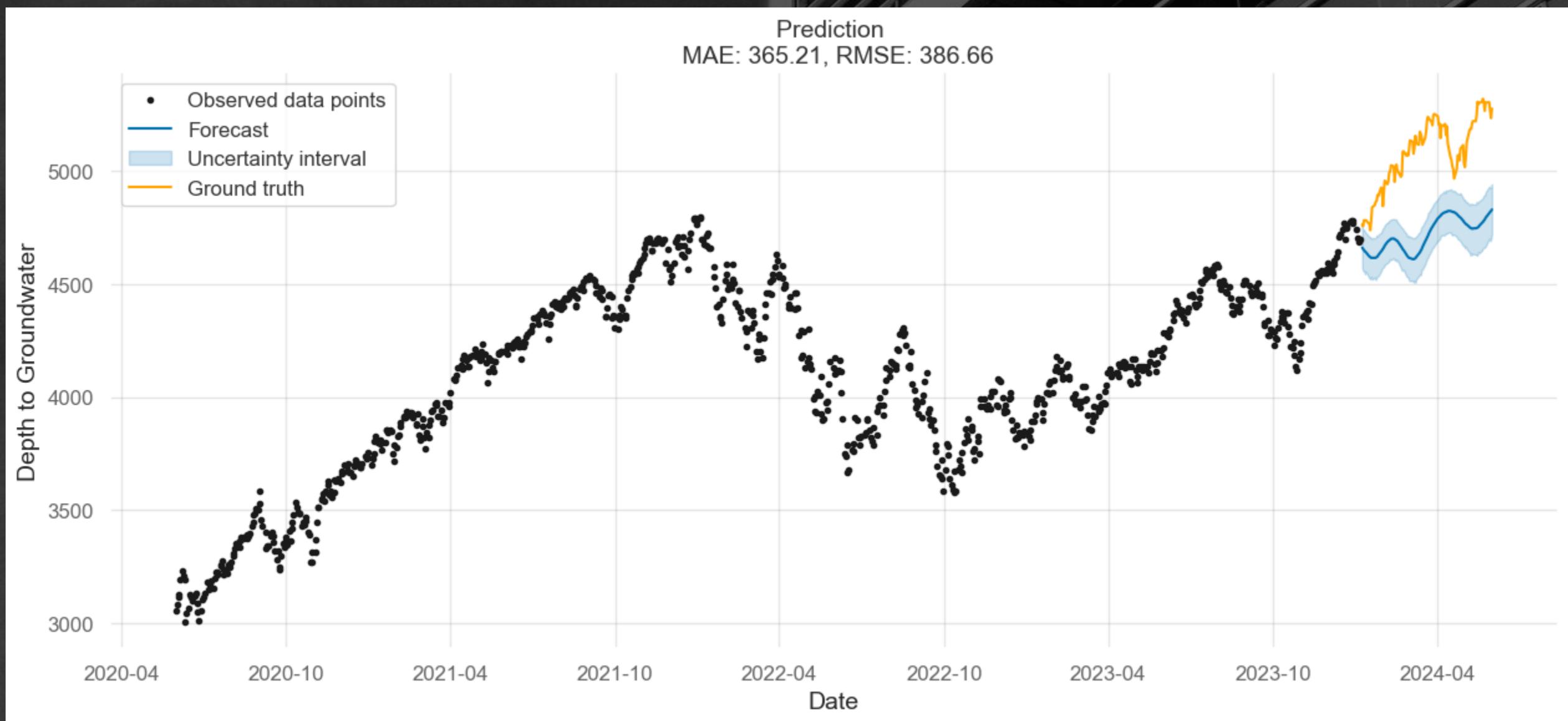
# Prophet

- Software Meta: Modelo tendencial
- Pretramiento de datos alto
- Captura estacionalidad, cambios tendencia festivos y minimiza errores



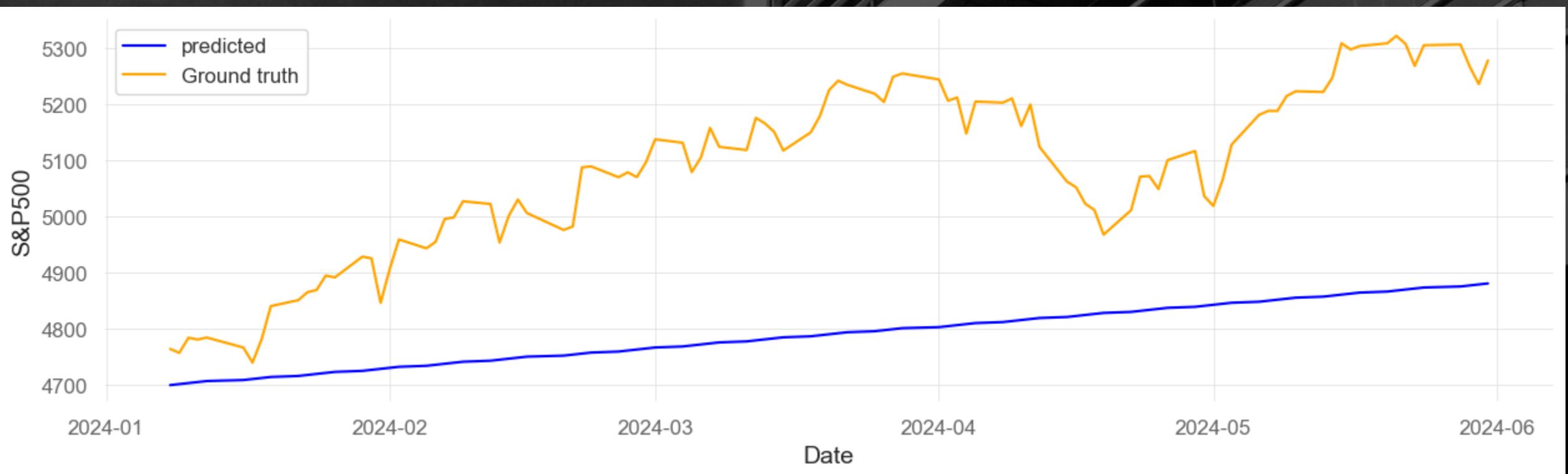
# Prophet + Box Cox

- Transformación para estabilizar la varianza
- Cambio con un parámetro calculado
- En este caso desplaza la predicción y empora el modelo



# ARIMA

- Auto Arima propone SARIMAX (0, 1, 0)
- Modelo no se adapta bien a la características del caso



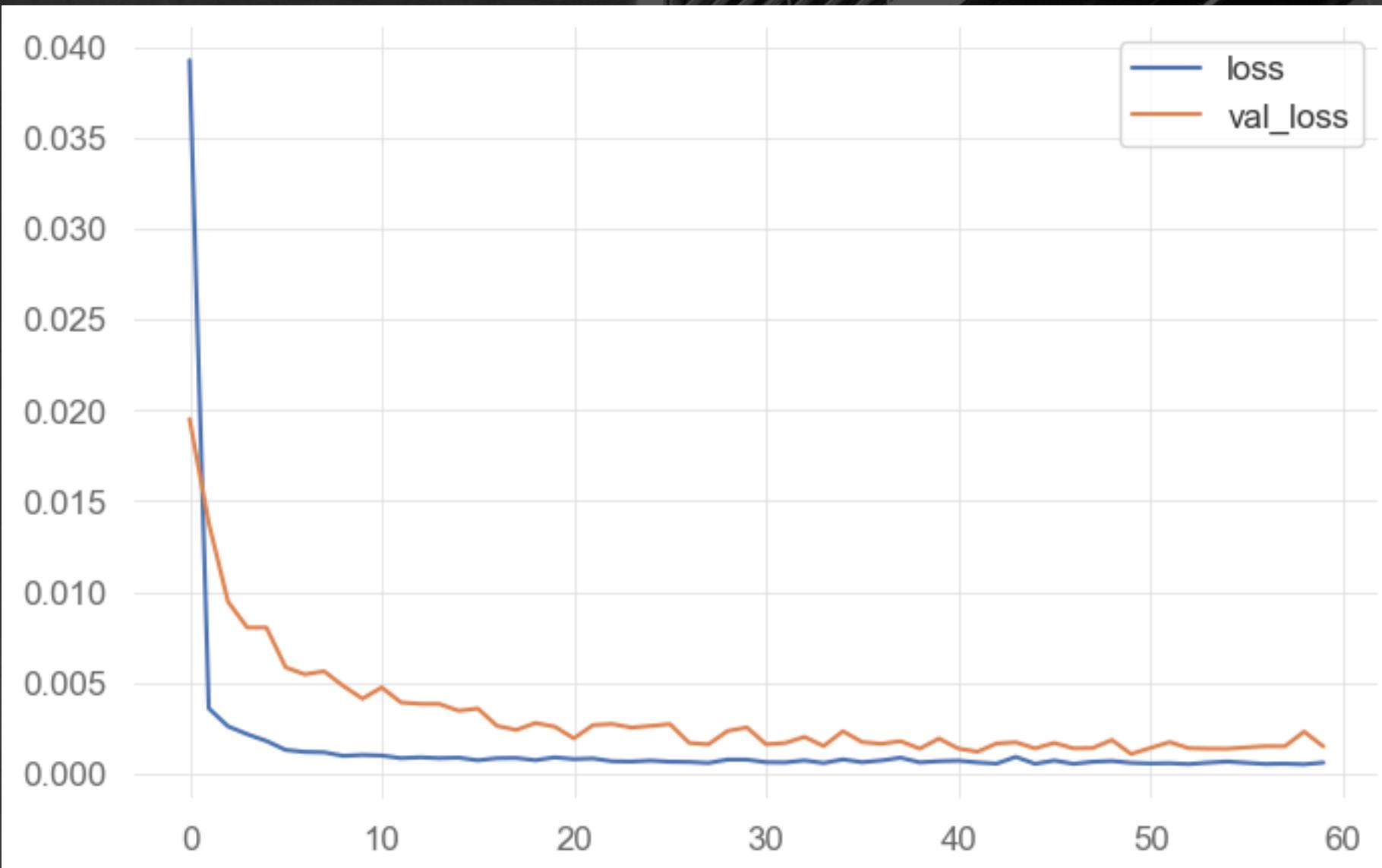
# LSTM

- o Modelo deep-learning centrado en Time Series
- o Resultados casi óptimos
- o MAE de 79 vs 55 del MA buscado
- o Buen ajuste a la tendencia



# LSTM

- o Se comprueba si hay overfitting
- o Curvas no se alejan a lo largo de los EPOCHS, validación del modelo

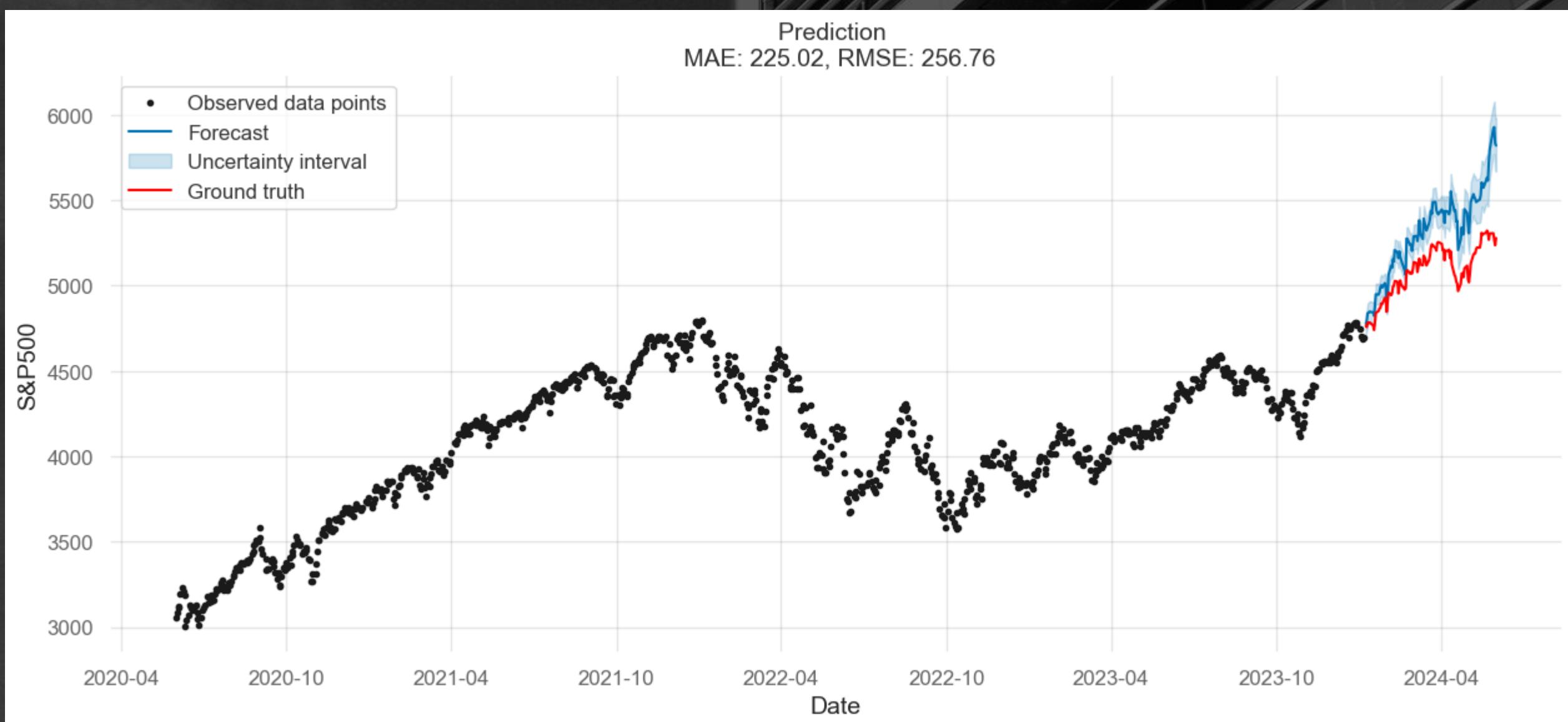


# MODELOS MULTIVARIANTES



# Prophet Multi

- Añadimos los parámetros del precio de close
- Predicciones muy altas: Sospecha modelo considerada a las features ruido.
- Añadiendo el volume empeorar el modelo



# CONCLUSIONES



Dificultad de  
predecir futuro



Features  
consideradas  
ruido



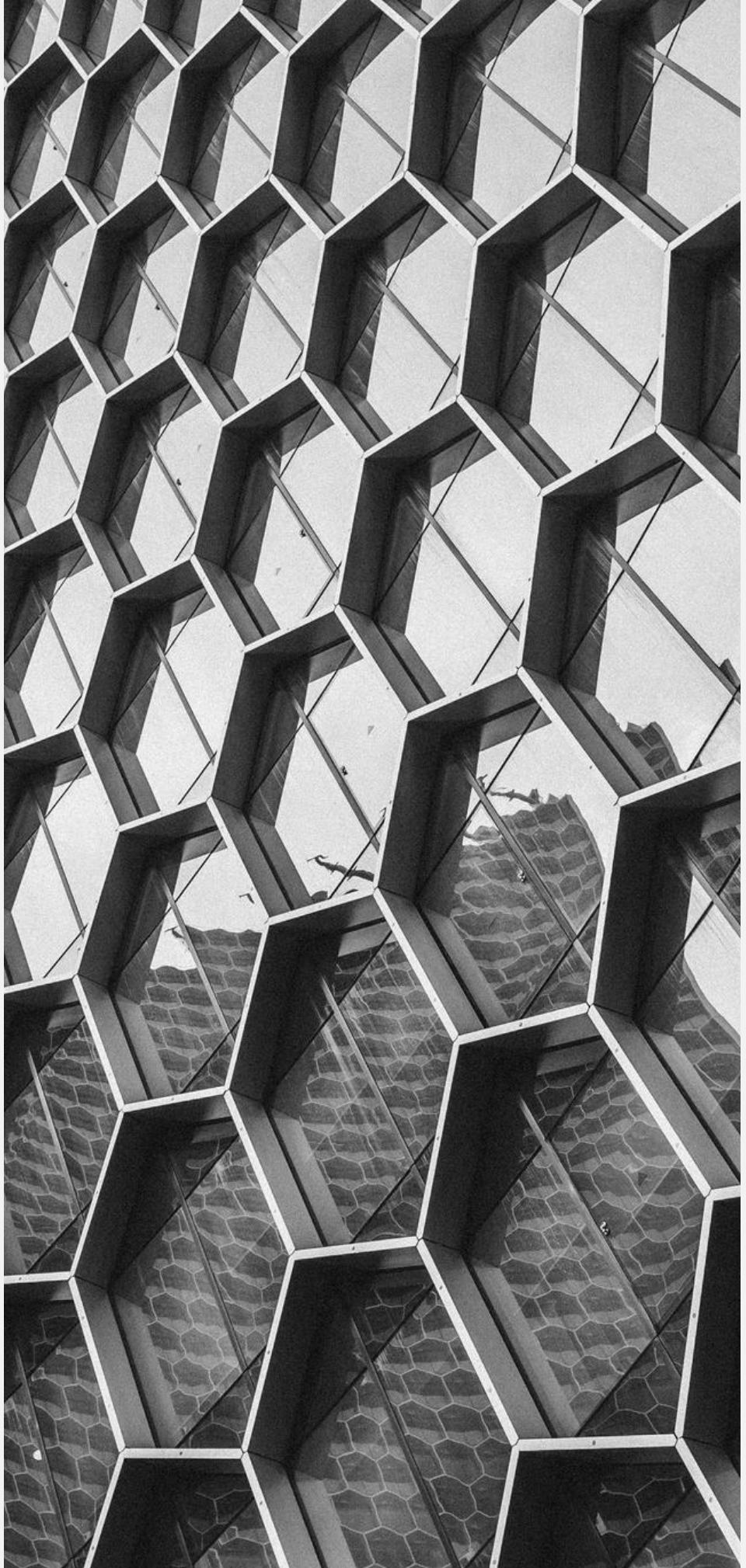
LSTM univariante  
Es el major  
modelo



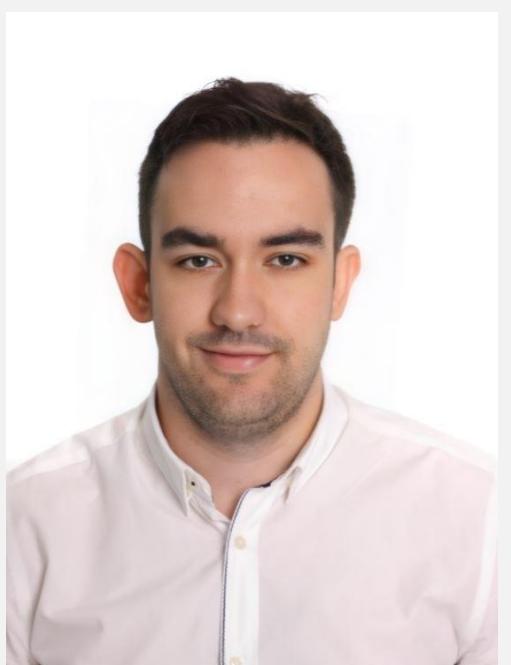
Prophet imita  
bien la tendencia  
pero el ajuste no  
es bueno



Modelo final es  
simple y se ajusta  
bien



**En caso de duda o consulta,  
póngase en contacto con:**



**Adrián Ruiz Donado**

- +34 647158470
- adrian.ruiz.donado@gmail.com
- <https://github.com/adriani94bil>
- Bilbao

**¡GRACIAS POR SU ATENCIÓN**