PREDICIENDO BITCOIN CON MACHINE LEARNING Y DEEP LEARNING



Xavier Roca Pérez

Programación de Redes Neuronales – Machine Learning

**Índice**

[**Motivación y descripción del problema** 3](#_Toc153448211)

[**Cuál es el objetivo inicial deseado?** 3](#_Toc153448212)

[**¿De dónde has extraído los datos?** 3](#_Toc153448213)

[**¿Cómo se han limpiado y preparado?** 4](#_Toc153448214)

[**Incluye algunos gráficos relevantes de la exploración de los datos** 5](#_Toc153448215)

[**¿Has aplicado “feature engineering”? En caso afirmativo, justifica las decisiones** 7](#_Toc153448216)

[**¿Qué algoritmos de ML has utilizado? ¿Por qué?** 7](#_Toc153448217)

[**¿Cómo has realizado el entrenamiento (se ha aplicado algún tipo de cross-validation, hyperparameter tuning, etc.)?** 8](#_Toc153448218)

[**¿Qué resultados has obtenido? Incluye algunos gráficos representativas** 10](#_Toc153448219)

[**Conclusiones** 11](#_Toc153448220)

[**¿Has logrado el objetivo que te habías propuesto? En caso negativo, ¿qué ha fallado?** 12](#_Toc153448221)

[**Describe brevemente cuáles podrían ser los siguientes pasos a realizar para mejorar el proyecto** 12](#_Toc153448222)

# **Motivación y descripción del problema**

Al ser un apasionado del mundo de las crypto y blockchain, me interesó aprender a como predecir bitcoin. Entonces busque diversos data sets para poder utilizar inteligencia artificial, como machine learning y deep learning para poder predecirlos mejor. Decidí utilizarlo de manera supervisada que es lo que hemos aprendido en el curso y más adelante, me gustaría utilizar reinforcement learning para que vaya aprendiendo solo, pero eso ya sería un extra.

# **Cuál es el objetivo inicial deseado?**

El objetivo deseado inicial es predecir el bitcoin, mirando diferentes maneras de hacerlo, he visto que mucha gente lo intenta predecir día a día, yo voy a intentarlo predecir cada dos horas.

# **¿De dónde has extraído los datos?**

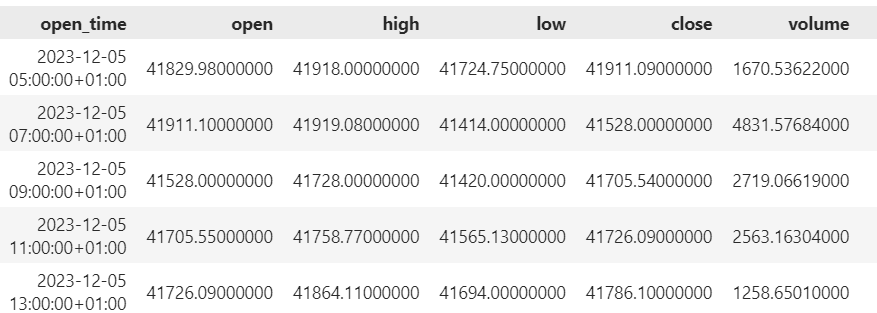
Los datos los he extraído de la api de binance:

'https://api.binance.com/'

He cogido los datos desde 2017 hasta el día 05-12-2023 de bitcoin, mas adelante me gustaría coger más datos como comparativas de diferentes crypto del top 10 con bitcoin para añadirlo al data set para que haya mas datos.

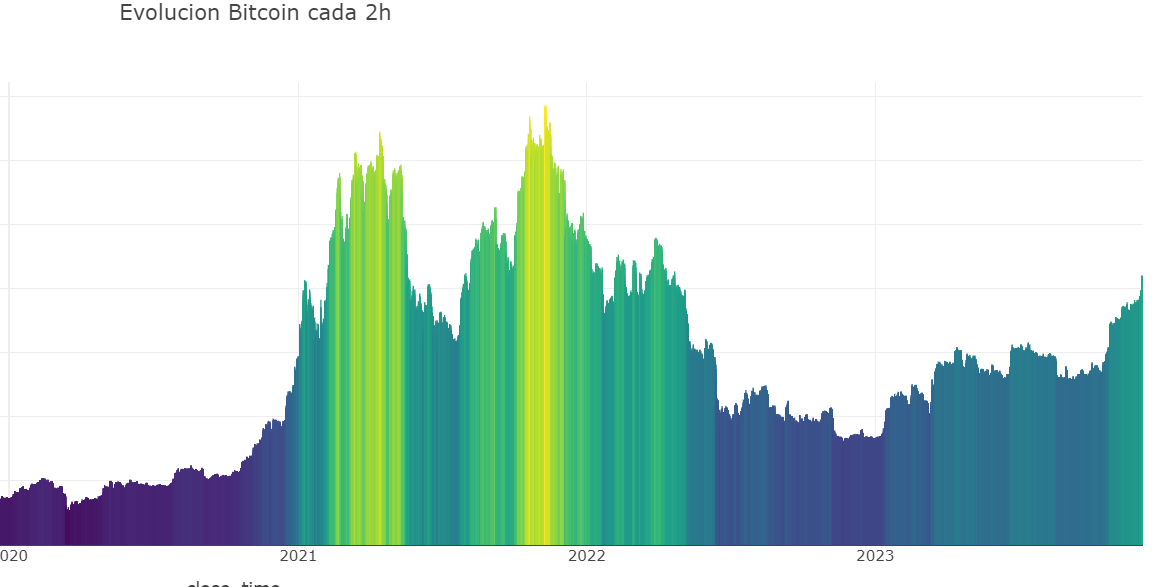
# **¿Cómo se han limpiado y preparado?**

Primero he cogido los datos de la api, después he añadido nombre a las columnas y he pasado la hora de numérica a hora europea en Madrid con el dia que representa. Por ultimo he añadido una columna target que sería lo que tendríamos que predecir que es el precio de close una hacia arriba.

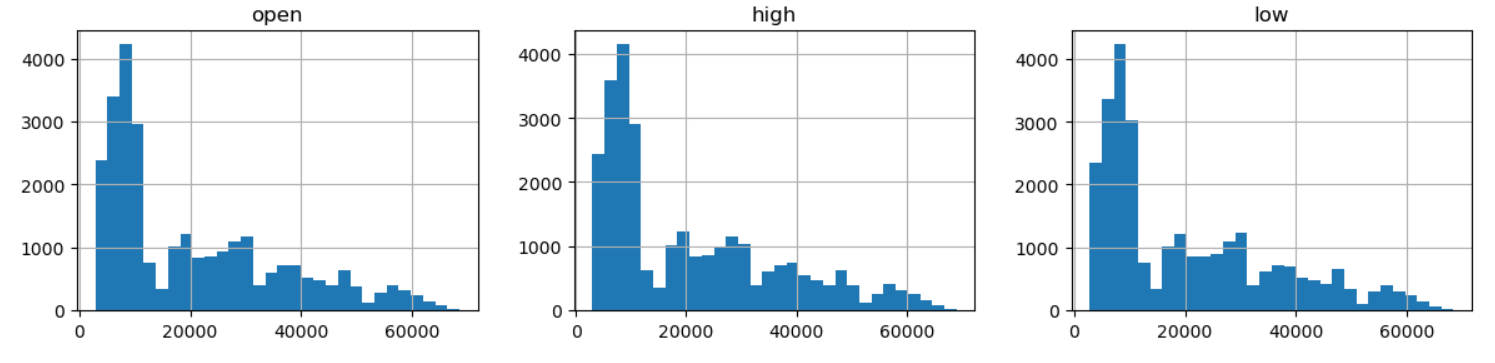


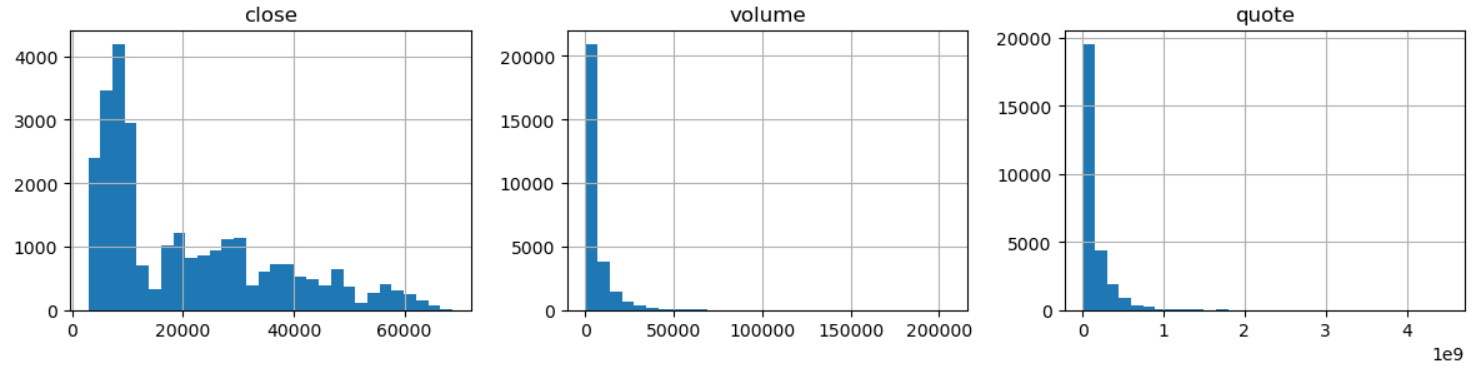


# **Incluye algunos gráficos relevantes de la exploración de los datos**

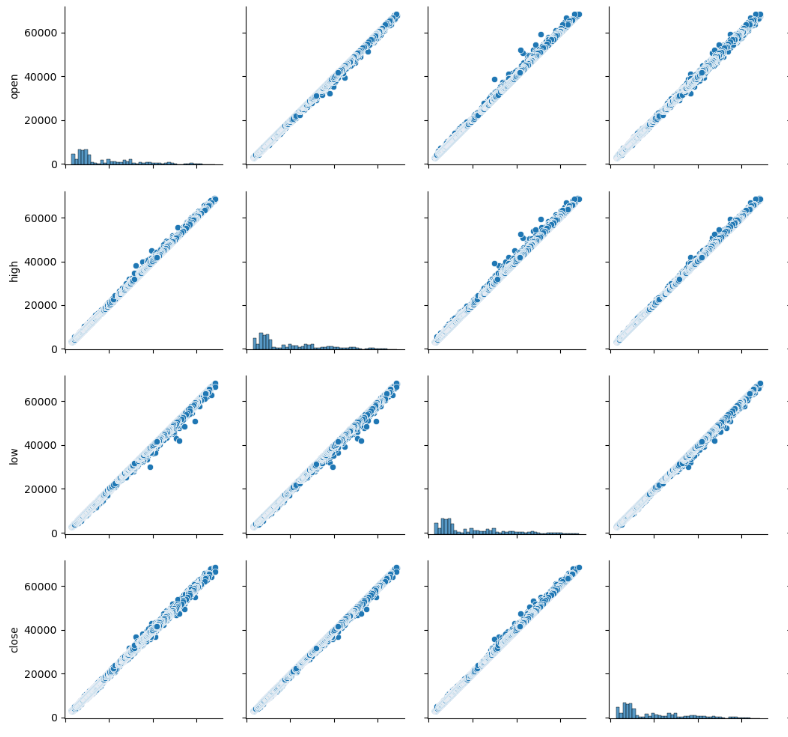


Numerical features distribution:



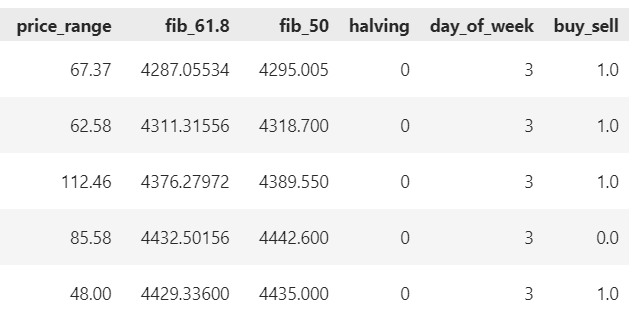


Pair plot:



# **¿Has aplicado “feature engineering”? En caso afirmativo, justifica las decisiones**

He hecho una predicción sin nuevos atributos y otro con nuevos atributos que creo que pueden ser interesantes.



# **¿Qué algoritmos de ML has utilizado? ¿Por qué?**

He utilizado varios algoritmos de ML, al nosotros querer predecir un numero numérico, he utilizado los algoritmos regressor para ello.

* Linear Regressor
* K-Neighbours
* Decision Tree
* Random Forest
* Gradiant Boosting Decision Tree
* SVR
* Elastic Net
* RNN
  + Simple
  + LSTM
  + Gru

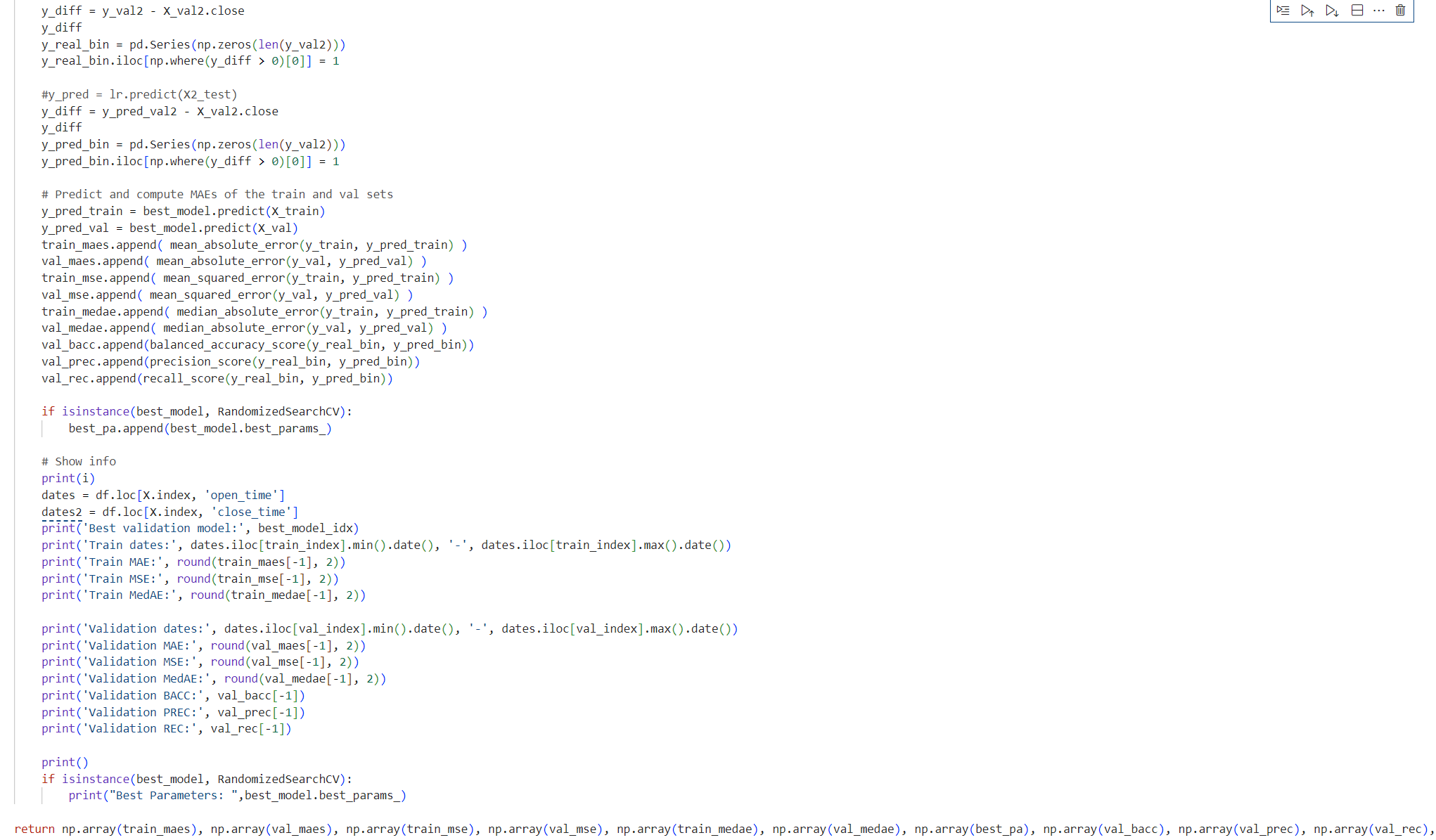
He ido probando varios algoritmos a ver cual daba mejor, las redes neuronales recurrentes no me han funcionado demasiado bien, ya que había un momento que no bajaba nada y no se quedaba en una buena cifra. De los otros algoritmos el que mejor resultado fue Linear Regressor y de ahí pensé que Elastic Net podría mejorarlo, pero al final el Linear Regressor es el que mejor resultado me ha dado.

# **¿Cómo has realizado el entrenamiento (se ha aplicado algún tipo de cross-validation, hyperparameter tuning, etc.)?**

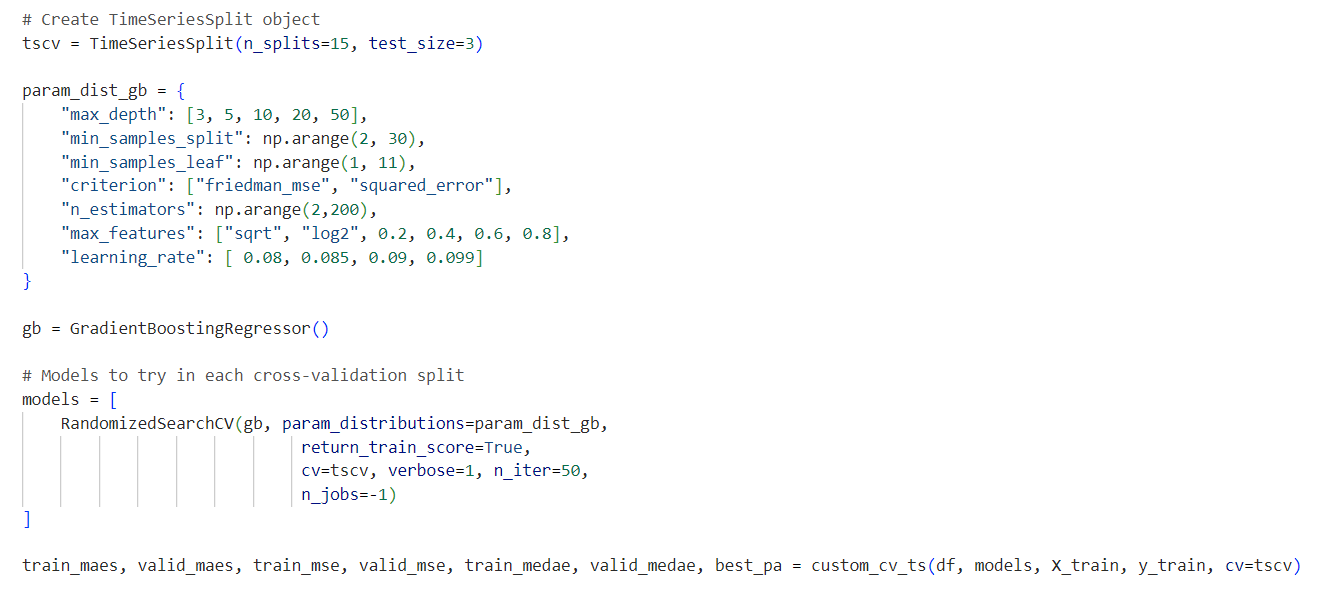
He utilizado un un cross-validation custom y en el hyperparameter tunning he utilizado el randomsearch.

Custom\_cv\_ts:



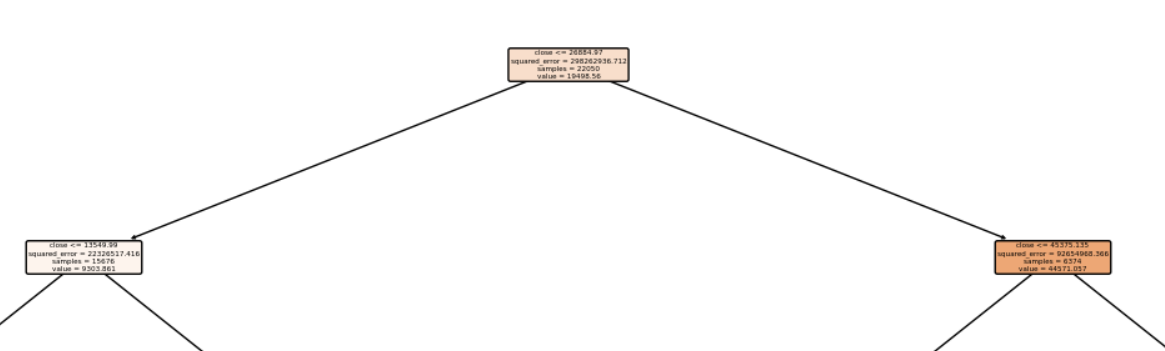


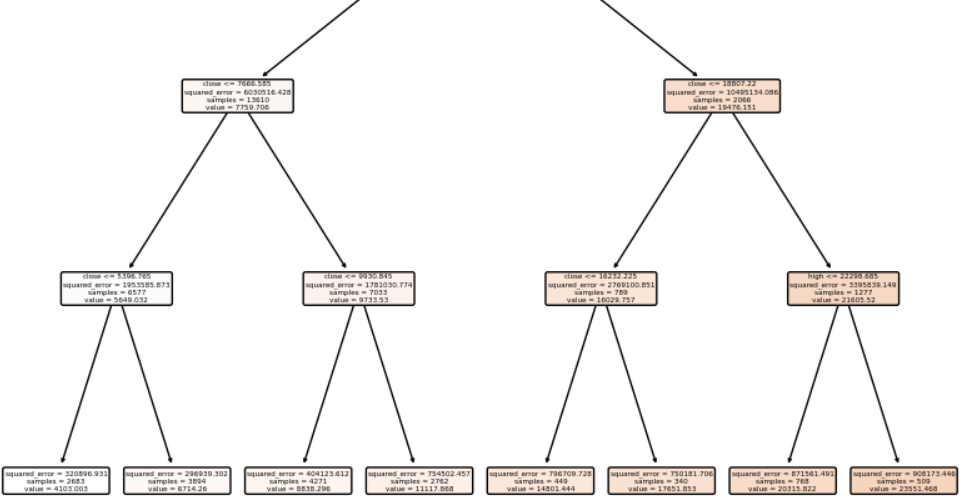
RandomSearch:

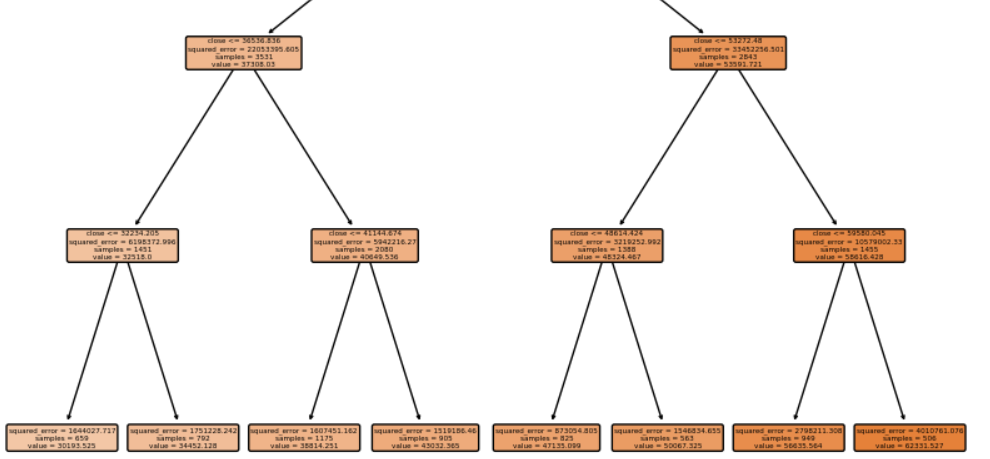


# **¿Qué resultados has obtenido? Incluye algunos gráficos representativas**

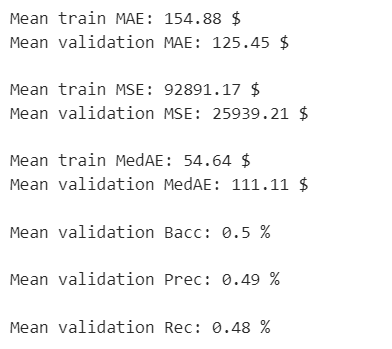
Decisión tree grafic:







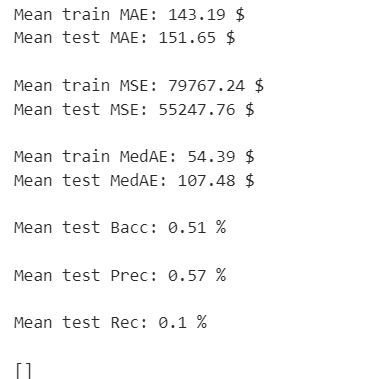
Best models metrics:



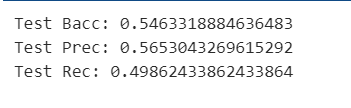
# **Conclusiones**

No entendía porque me aba el recall tan bajo, así que lo probé de otra manera

Best models:



Linear Regression: fue el modelo que mejor me fue



# **¿Has logrado el objetivo que te habías propuesto? En caso negativo, ¿qué ha fallado?**

He conseguido a medias el objetivo que tenía planteado, ya que he podido ver cómo hacerlo, las métricas balance accuracy y precisión no están mal pero el recall haciendo los tests es muy bajo, entonces hice otras pruebas, y vi que el recall ya me subía al 50%, entonces creo que es una buena métrica, aunque para poder invertir con este modelo, querría que fuera más del 60% o 70% de recall, aun así podría hacer un backtesting para probarlo y ver resultados. Otra cosa que me he dado cuenta es que lo que obtengo en mi mejor modelo con la mejor métrica no mejora de mucho lo obtenido con la baseline, entonces nos damos cuenta que no es un gran modelo.

# **Describe brevemente cuáles podrían ser los siguientes pasos a realizar para mejorar el proyecto**

Los siguientes pasos a solucionar son, intentar descubrir una red neuronal que pueda funcionar y ampliar el data set con más datos para que le sea más fácil predecir el precio cada dos horas, hacer backtesting para probar si con el modelo se puede ganar dinero.