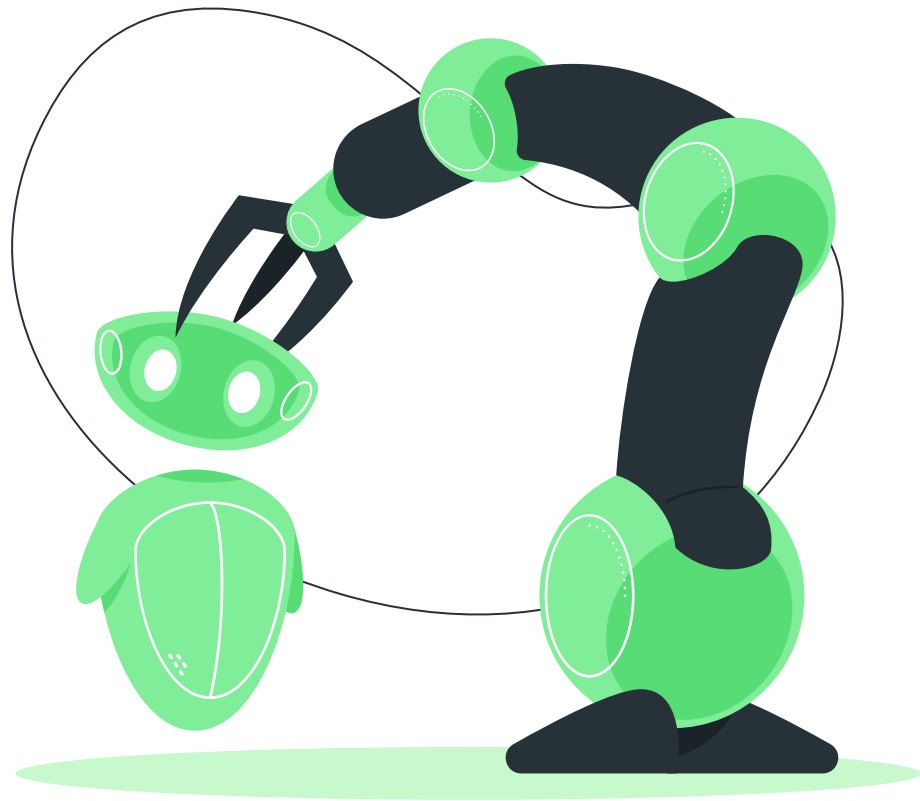


Reto 3:

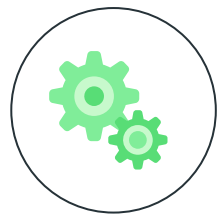
Script Kiddies



Cristina Guerrero Cano
Sergio Quintero Castellanos
Ismael Dotor Vera
Pablo Ribas Borrego



Limpieza



Outliers

Método IQR

Borrar

Sustituir por null values



Null values

Promedio de la
variación de los
últimos 10 días

Open

Close

Bayesian posterior
consensus
distribution

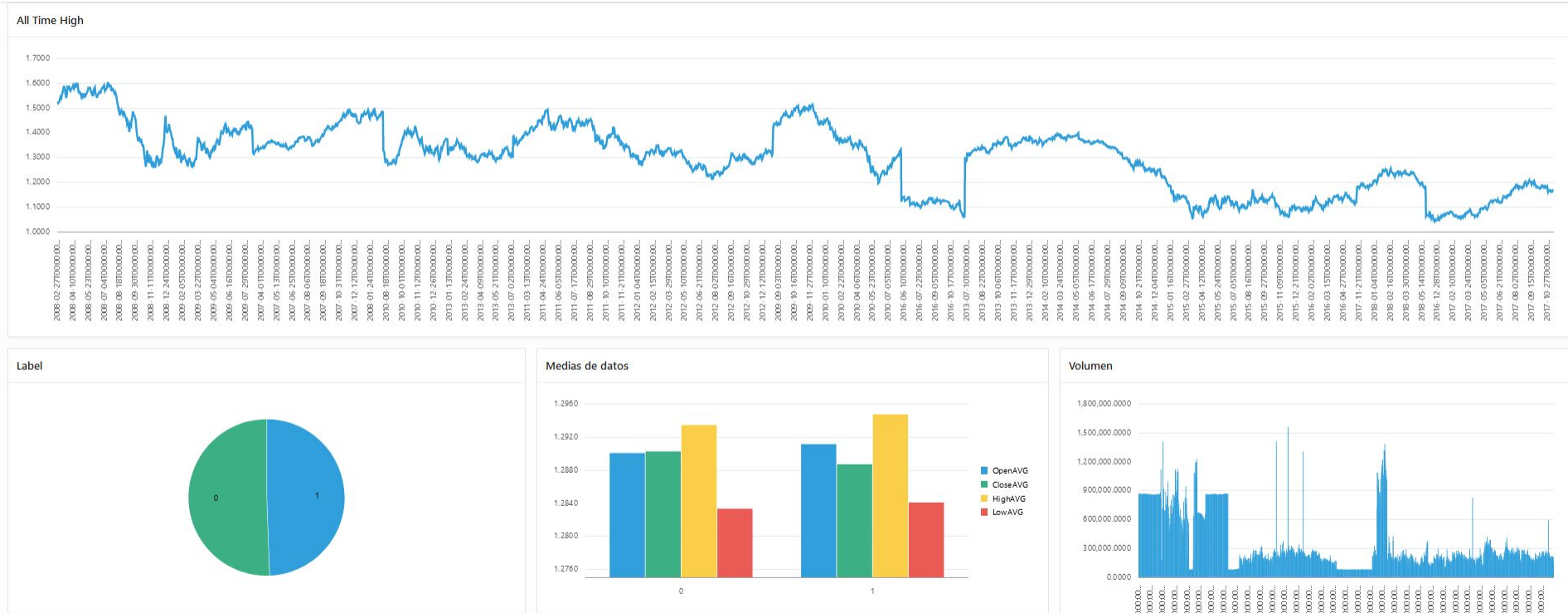
High

Low

Volume

Análisis exploratorio

Dashboard



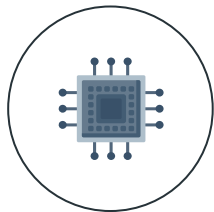
[Link a la app de Apex](#)

Modelos probados



SVM

Máquinas de soporte vectorial
F1 Score: 0,4710



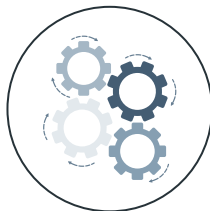
Ada Boost

Combinación de clasificadores
F1 Score: 0,477



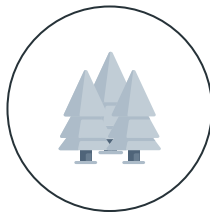
KNN

Vecinos más cercanos
F1 Score: 0,5015



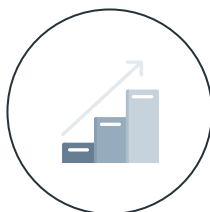
Naive-Bayes

Modelo Bayesiano ingenuo
F1 Score: 0,454



Random Forest

Bosque de árboles de decisión
F1 Score: 0,467



Gradient Boosting

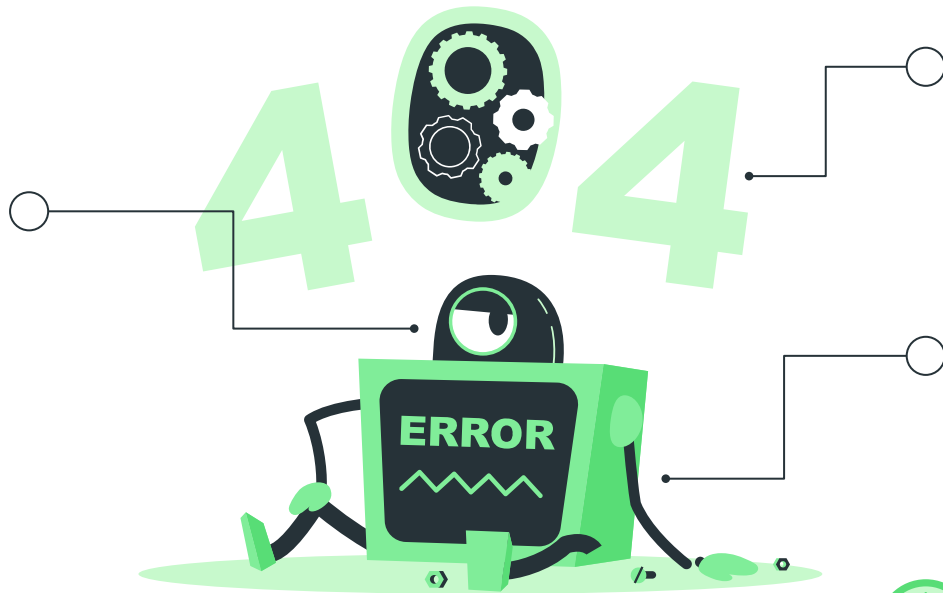
Conjunto de predictores
F1 Score: 0,4652

Error en el enfoque

Fechas

Para el modelo se debe tener en cuenta la **temporalidad de los datos**

De
07/03/2007
a
23/05/2018



ARIMA

Primer modelo.

Añadir días faltantes como nulos y rellenarlos

XGBoost

Modelo final para la predicción.

ANTERIORES MODELOS  TIEMPO

XGBoost – modelo final

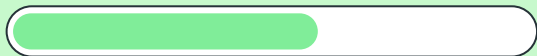
Suponemos que: *Label* 0/1 indica la fluctuación del precio de mercado

No queremos utilizar datos futuros reales
(Look-Ahead Bias)

Predecimos los valores basándonos en el pasado

Establecer *Label* 0/1 en función de predicción *Close*

51%

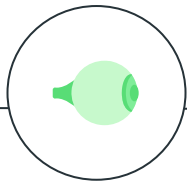


F1 Macro



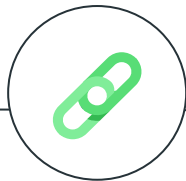
Entrenamiento y validación

Paso 1



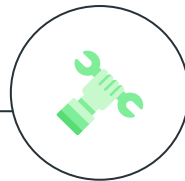
Visualización Serie Temporal

Paso 2



Buscar relaciones características/objetivo

Paso 3



Creación del modelo

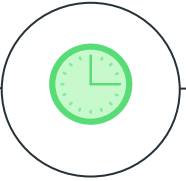
**AJUSTE
PARÁMETROS**

Paso 4

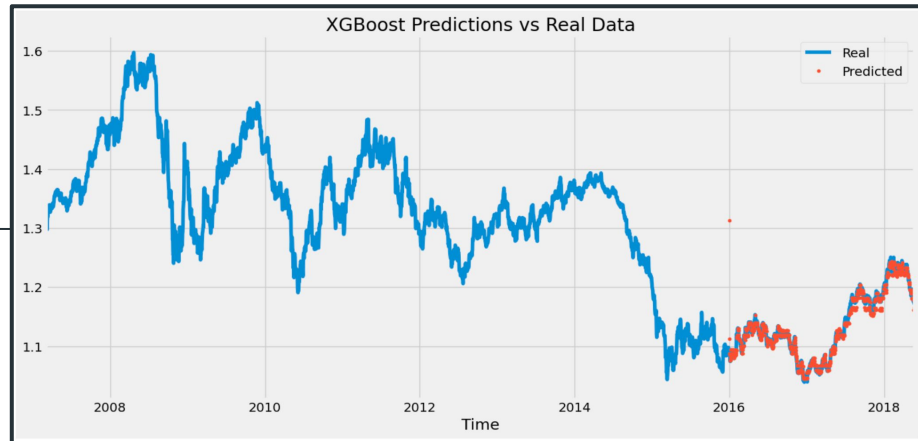


Evaluar importancia características

Paso 5



Predicciones para futuro



Resultados y conclusiones

- Enfoque inicial incorrecto
 - No se tenía en cuenta temporalidad
- Importancia de la temporalidad de los datos.
 - Necesidad de tener en cuenta los valores diarios.
 - Modelo compatible con series temporales → XGBoost.
- Desarrollo XGBoost
 - ¡Evitar LOOK-AHEAD BIAS!
 - Evitar sobreajuste
 - Predicción de valor *Close* futuro para 0/1 en función de *de/*crecimiento.
- ¿Significado de la columna a predecir *label*? ¿De qué variables depende?
 - Dificultad a la hora de realizar la predicción.
 - Falta de información.



¡Muchas gracias!

¿Alguna pregunta? ¡Contáctanos!

pablo.ribas.borrego@gmail.com