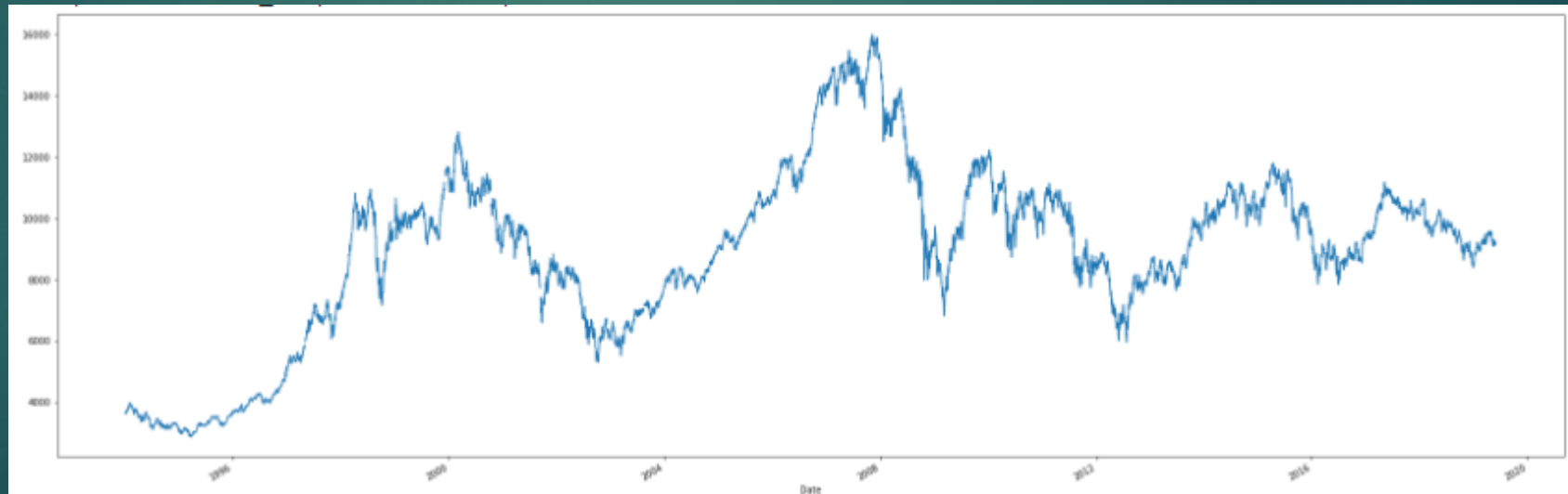


Reto de Ciencia **de** Datos

Alvaro Ramírez Sixtos

Exploración de los datos

- ▶ Se encontraron valores faltantes (133).
- ▶ No se encontraron valores atípicos.
- ▶ Las variables tenían casi similar el comportamiento en el tiempo.



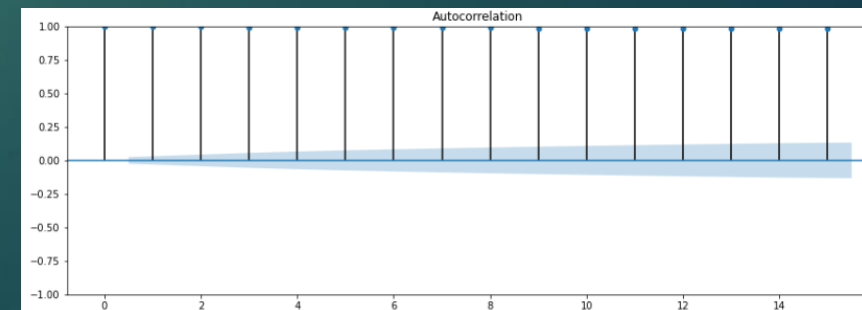
Preprocesamiento y Modelos de clasificación

- ▶ Se realizó interpolación de los datos para completar los valores faltantes.
- ▶ Al tener una baja correlación entre las features y la variable predictora, un algoritmo de clasificación no sería de gran ayuda. Se comprobó realizando modelos de clasificación utilizando SVM o RandomForest y el resultado fue el esperado. La precisión era prácticamente aleatoria (48% de precisión).

	precision	recall	f1-score	support
0	0.48	1.00	0.65	631
1	0.00	0.00	0.00	680
accuracy			0.48	1311
macro avg	0.24	0.50	0.32	1311
weighted avg	0.23	0.48	0.31	1311

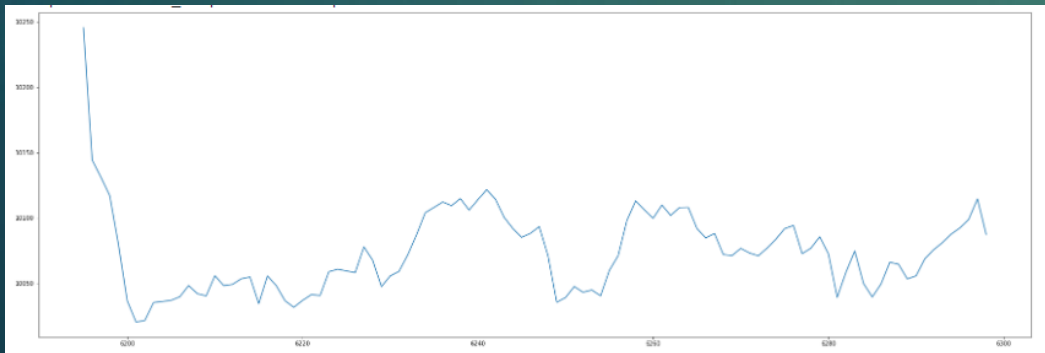
Modelado de Serie Temporal

- ▶ Se optó por generar un análisis de serie temporal para predecir el Close que tendría la serie después de 3 días.
- ▶ Se graficó la autocorrelación de la variable Close y se encontró que los valores anteriores podrían servir para predecir un valor futuro.
- ▶ Se generaron variables adicionales, como las diferencias, medias y desviaciones estándar de la variable Close con respecto a 1, 3 y 7 días. Además variables categóricas como Día de la semana y Mes.
- ▶ Se entrenaron modelos de RandomForestRegressor y XGBRegressor, a éstos se le modificaron hiper parámetros para encontrar el mejor modelo.

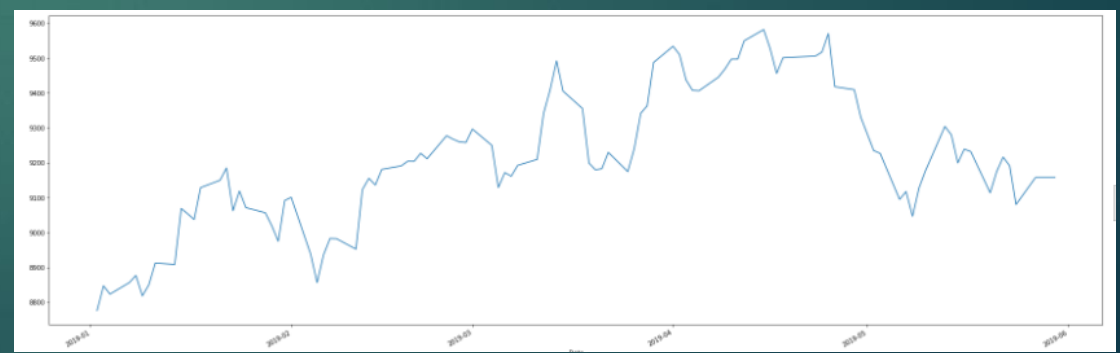


Resultados

- ▶ El modelo con el mejor desempeño fue el RandomForestRegressor, este modelo predice el valor de la variable Close para tres días después dada la información del día actual.
- ▶ Para generar la variable Target solamente se comprobó que el valor predicho con el valor actual de Close.
- ▶ Al probar el performance de nuestro modelo, se obtuvo un 57 % de precisión.



Predicción



Real