



Instituto Politécnico Nacional

Escuela Superior de Cómputo



LICENCIATURA EN CIENCIA DE DATOS

UNIDAD DE APRENDIZAJE

ANÁLISIS DE SERIES DE TIEMPO

TÍTULO DEL TRABAJO

Práctica 7. Pronósticos modelo ARIMA

ALUMNOS

Aguilar Ramírez Carlos Francisco

Arista Romero Juan Ismael

Jiménez Flores Luis Arturo

Vázquez Martín Marlene Gabriela.

PROFESOR (A)

Daniel Jiménez Alcantar

GRUPO

6AV1

Fecha de entrega: 11 de mayo de 2025

Introducción

En esta práctica se desarrolló un modelo de pronóstico utilizando técnicas de series de tiempo aplicadas a los precios ajustados de las acciones de Netflix (NFLX). El objetivo fue construir un modelo ARIMA siguiendo la metodología Box-Jenkins, a partir de datos diarios comprendidos entre mayo de 2002 y diciembre de 2022, y luego validar su capacidad predictiva utilizando los datos reales del año 2023. Esta serie representa un caso real del comportamiento financiero de una empresa tecnológica, lo cual añade complejidad debido a su alta volatilidad y eventos externos que afectan su dinámica. El enfoque de este análisis se centra en identificar, ajustar y validar un modelo ARIMA que permita evaluar su desempeño al predecir valores futuros de la acción.

Problemática

Los precios de las acciones fluctúan constantemente debido a múltiples factores, tanto internos como externos a la empresa. Esta naturaleza volátil complica el desarrollo de modelos de predicción que mantengan una precisión aceptable en el tiempo. En el caso específico de Netflix, la evolución del precio ajustado de sus acciones a lo largo de más de dos décadas presenta fuertes tendencias, rupturas abruptas y periodos de crecimiento o caída impredecibles. Ante este panorama, surge la necesidad de aplicar un modelo que permita capturar la estructura temporal de la serie y realizar pronósticos a corto plazo. El reto consiste en validar si un modelo ARIMA puede ofrecer resultados razonablemente cercanos a los valores reales, incluso en un entorno tan dinámico como el del mercado bursátil.

Modelo estadístico

El modelo estadístico central utilizado en esta práctica es el *AutoRegressive Integrated Moving Average* (ARIMA). Un modelo ARIMA se define por tres órdenes: (p, d, q).

- p: Orden de la parte Autorregresiva (AR), que modela la dependencia de un valor actual con valores pasados.
- d: Orden de la Diferenciación (Integrated - I), que indica cuántas veces se debe diferenciar la serie para hacerla estacionaria.
- q: Orden de la parte de Medias Móviles (MA), que modela la dependencia de un valor actual con errores de predicción pasados.

La selección de estos órdenes (p, d, q) es el objetivo principal de la inferencia de parámetros mediante la metodología Box-Jenkins.

Modelo computacional

El modelo computacional se implementa en Python utilizando un conjunto de librerías especializadas:

Pandas: Para la manipulación y carga de datos (DataFrames y Series).

NumPy: Para operaciones numéricas.

Matplotlib y Seaborn: Para la visualización de datos (gráficos de series, ACF/PACF, residuos, etc.).

Statsmodels: Librería clave que proporciona las herramientas para el análisis de series de tiempo, incluyendo:

- Prueba de Dickey-Fuller Aumentada (adfuller) para estacionariedad.
- Funciones para graficar ACF y PACF (plot_acf, plot_pacf).
- Implementación del modelo ARIMA (ARIMA).
- Descomposición estacional (seasonal_decompose).
- Pruebas de diagnóstico de residuos (ej., Ljung-Box a través de acorr_ljungbox).

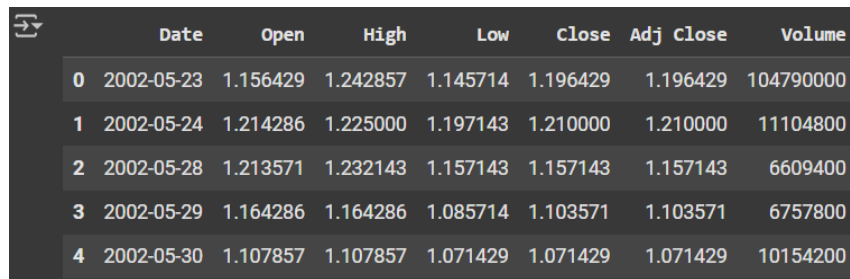
Scipy: Utilizada para encontrar picos y valles (find_peaks) en el EDA.

Sklearn: Utilizada para métricas de evaluación como el Error Cuadrático Medio (mean_squared_error).

El código está estructurado en funciones para mejorar la modularidad y reutilización, cubriendo desde la carga de datos hasta la evaluación del modelo.

Metodología

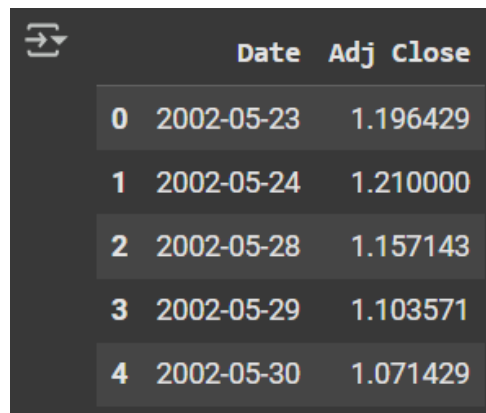
El desarrollo de esta práctica se llevó a cabo utilizando la metodología Box-Jenkins, la cual proporciona una estructura sistemática para la construcción, ajuste y validación de modelos ARIMA. Sin embargo, el proceso comenzó con la carga y limpieza del conjunto de datos, conformado por los precios ajustados diarios de las acciones de Netflix (Adj Close) desde mayo de 2002 hasta diciembre de 2023.



	Date	Open	High	Low	Close	Adj Close	Volume
0	2002-05-23	1.156429	1.242857	1.145714	1.196429	1.196429	104790000
1	2002-05-24	1.214286	1.225000	1.197143	1.210000	1.210000	11104800
2	2002-05-28	1.213571	1.232143	1.157143	1.157143	1.157143	6609400
3	2002-05-29	1.164286	1.164286	1.085714	1.103571	1.103571	6757800
4	2002-05-30	1.107857	1.107857	1.071429	1.071429	1.071429	10154200

Figura 1. Carga y preparación de los datos

Para facilitar el análisis, se conservaron únicamente las columnas esenciales: la fecha de registro y el precio ajustado, eliminando así cualquier información irrelevante para el modelado.



	Date	Adj	Close
0	2002-05-23	1.196429	
1	2002-05-24	1.210000	
2	2002-05-28	1.157143	
3	2002-05-29	1.103571	
4	2002-05-30	1.071429	

Figura 2. Limpieza y transformación

Posteriormente, se identificó el comportamiento general de la serie mediante visualización gráfica, lo cual reveló una fuerte tendencia no estacionaria, con fases de crecimiento acelerado y caídas abruptas.

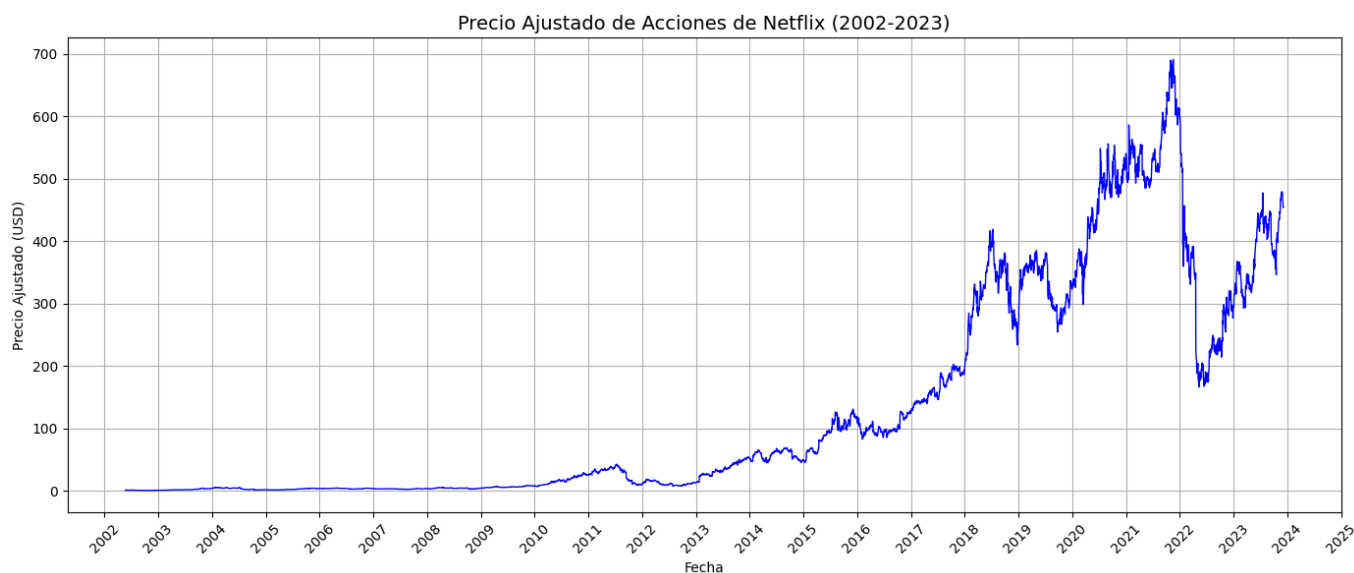


Figura 3. Visualización de la serie de tiempo.

1) Box - Jenkins: Identificar el modelo

Para confirmar esta observación de forma estadística, se aplicó la prueba de Dickey-Fuller, la cual arrojó un p-valor superior al 0.05, lo que indicó que la serie no era estacionaria. En consecuencia, se aplicó una primera diferenciación, tras la cual se repitió la prueba obteniendo un p-valor prácticamente nulo, lo que confirmó que la serie ya era estacionaria.

Con la serie diferenciada, se procedió al análisis de la función de autocorrelación (FAC) y autocorrelación parcial (FACP). Los resultados mostraron que solo el primer rezago tenía correlación significativa, tanto en FAC como en FACP, lo cual sugirió que un modelo ARIMA (0,1,1) podría ser adecuado.

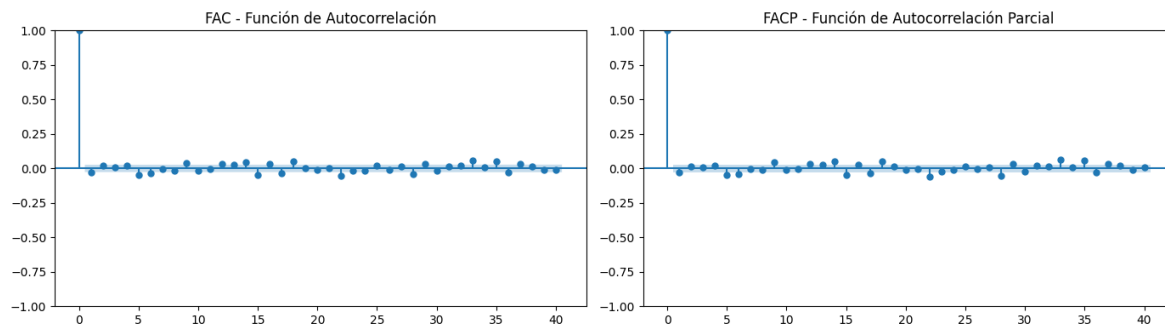


Figura 4. Funciones de autocorrelación.

2) Box Jenkins: Estimación de parámetros

A continuación, se realizó el ajuste del modelo utilizando únicamente los datos de entrenamiento, definidos desde 2002 hasta el 31 de diciembre de 2022. Se estimaron los parámetros y se obtuvieron los residuos del modelo, los cuales fueron analizados tanto visualmente como mediante histogramas. Los residuos se distribuyeron en torno a cero sin mostrar patrones visibles, lo cual sugiere que el modelo fue capaz de capturar la dinámica principal de la serie.

SARIMAX Results						
Dep. Variable:	Adj Close	No. Observations: 5189				
Model:	ARIMA(0, 1, 1)	Log Likelihood -16328.097				
Date:	Wed, 07 May 2025	AIC 32660.193				
Time:	00:58:31	BIC 32673.302				
Sample:	0	HQIC 32664.779				
- 5189						
Covariance Type: opg						
	coef	std err	z	P> z	[0.025 0.975]	
ma.L1	-0.0337	0.007	-5.170	0.000	-0.046	-0.021
sigma2	31.7115	0.091	347.664	0.000	31.533	31.890
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB): 1825130.05				
Prob(Q):	0.95	Prob(JB): 0.00				
Heteroskedasticity (H):	5845.88	Skew: -2.76				
Prob(H) (two-sided):	0.00	Kurtosis: 94.72				

Figura 5. Parámetros estimados para el modelo ARIMA

Luego de ajustar el modelo ARIMA (0,1,1) sobre la serie diferenciada, se obtuvo un coeficiente para el término MA (1) de -0.0337 , lo cual indica que el modelo considera muy poca influencia del error del periodo anterior en la predicción actual. Además, la varianza de los errores (σ^2) fue de aproximadamente 31.71. Ambos parámetros resultaron estadísticamente significativos, lo que valida su inclusión en el modelo. A pesar de que la serie presenta una distribución no normal en los residuos (algo habitual en series financieras), el ajuste general del modelo resulta adecuado para fines de pronóstico de corto plazo.

3) Box Jenkins: Uso del modelo y validación.

Finalmente, se utilizaron los datos del año 2023 (previamente reservados como conjunto de validación) para generar pronósticos y compararlos con los valores reales. Esta comparación se realizó de forma gráfica y cuantitativa,

utilizando métricas como MAE, RMSE y MAPE, lo que permitió evaluar la precisión del modelo de forma objetiva.



Figura 6. Predicción con ARIMA frente a datos reales.

Resultados del modelo:

- MAE (Error Absoluto Medio): 89.23
- RMSE (Raíz del Error Cuadrático Medio): 102.80
- MAPE (Error Porcentual Medio Absoluto): 21.84%

Figura 7. Validación del modelo ARIMA

Derivado de esto, podemos interpretar que, en promedio, el modelo se desvía unos 89 dólares del valor real por día. Además, el error porcentual promedio del pronóstico respecto a los valores reales fue de aproximadamente 22%, lo cual es aceptable para una serie financiera con alta volatilidad.

Conclusiones

Aguilar Ramírez Carlos Francisco

La implementación del modelo ARIMA bajo la metodología Box-Jenkins fue una oportunidad valiosa para aplicar de manera estructurada los conceptos teóricos en un caso real. Trabajar con los precios ajustados de las acciones de Netflix me permitió experimentar de forma práctica todo el ciclo de modelado: desde la detección inicial de no estacionariedad mediante la prueba de Dickey-Fuller, hasta la selección del modelo más adecuado apoyándome en los gráficos de autocorrelación.

Elegimos un modelo ARIMA (0,1,1), el cual, si bien es simple, resultó razonablemente eficaz al capturar parte de la dinámica de corto plazo de la serie. Lo más interesante fue constatar cómo, pese a tener un buen ajuste estadístico y residuos aceptables, las predicciones no logran anticipar los cambios bruscos del mercado, lo que me permitió reflexionar sobre las limitaciones de ARIMA cuando se enfrenta a series financieras de alta volatilidad.

Esta práctica reforzó no solo mis habilidades técnicas, sino también la importancia de acompañar los resultados con análisis crítico y visual para detectar los límites reales del modelo.

Arista Romero Juan Ismael

El proceso de construir un modelo ARIMA con datos reales del mercado bursátil fue, sin duda, un reto diferente al que normalmente se presenta en ejercicios académicos. Utilizar los precios diarios de Netflix implicó enfrentarse a una serie con alta variabilidad y sin patrones cíclicos claramente definidos. Esto me hizo comprender la utilidad del componente “I” de ARIMA, ya que fue necesario diferenciar la serie para lograr estacionariedad.

Uno de los aprendizajes más importantes para mí fue la utilidad de FAC y FACP para orientar la elección del modelo. Aunque finalmente se optó por ARIMA(0,1,1), me di cuenta de lo fundamental que es observar con atención la estructura de la serie, ya que los datos financieros, a diferencia de los meteorológicos, rara vez presentan regularidad. Esta práctica no solo me permitió aplicar los conceptos del curso, sino también afinar mi criterio sobre cuándo y cómo usar un modelo estadístico, recordando siempre que el valor de un modelo no está solo en las métricas, sino también en su interpretación en contexto.

Jiménez Flores Luis Arturo

Durante esta práctica analizamos los precios ajustados de Netflix utilizando la metodología Box-Jenkins para construir un modelo ARIMA. Lo que más me ayudó a consolidar mi aprendizaje fue seguir el flujo completo del proceso, desde el

análisis inicial de estacionariedad con la prueba ADF hasta el ajuste y validación del modelo. Pudimos comprobar que la serie no era estacionaria, lo que nos llevó a aplicar una diferenciación para luego estimar y validar un modelo ARIMA (0,1,1).

Algo que destaco fue el análisis de residuos, pues, aunque estadísticamente parecían adecuados, al comparar las predicciones con los datos reales, se observó que el modelo no era capaz de anticipar movimientos bruscos del mercado. Este contraste entre resultados técnicos y la realidad del comportamiento financiero me hizo valorar más la necesidad de validar visual y cuantitativamente un modelo. Aunque ARIMA fue útil, también me hizo pensar en otras herramientas más robustas para series con alta volatilidad como esta.

Vazquez Martin Marlene Gabriela

Aplicar la metodología Box-Jenkins a una serie financiera como la de los precios de Netflix fue una experiencia diferente a otros ejercicios más regulares. Fue interesante ver cómo el modelo ARIMA (0,1,1), aunque técnicamente bien ajustado, tiene limitaciones frente a datos tan inestables. Aprendí la importancia de analizar si una serie es estacionaria, y cómo la diferenciación es un paso clave para estabilizarla y hacerla apta para modelado.

Una lección que me llevo es que no basta con que los residuos parezcan ruido blanco o que las métricas de error sean razonables; también hay que observar si el modelo logra reproducir la lógica de los datos reales. En este caso, al tratarse de una acción tan sensible a factores externos, entendí que modelos más sofisticados o híbridos podrían dar mejores resultados. Sin embargo, esta práctica me permitió entender muy bien los fundamentos de ARIMA y la utilidad del enfoque paso a paso de Box-Jenkins para abordar cualquier serie de tiempo.