

Pronóstico del precio semanal para Bancolombia (CIB) por medio de ChatGPT 3.5

Juan Sebastian Moncada Aguilar

Prompt



Por medio de un script de Python extrae el precio de Bancolombia desde Yahoo Finance con periodicidad semanal durante los últimos 5 años y realiza un modelo de pronóstico del precio para cada una de las siguientes 4 semanas.

Resultados ChatGPT

Para lograr esto, necesitarás utilizar algunas bibliotecas de Python, como `yfinance` para obtener los datos de Yahoo Finance y `pandas` para el análisis y manipulación de los datos, así como una librería de machine learning, por ejemplo `scikit-learn`, para construir el modelo de pronóstico. Asumiré que ya tienes estas bibliotecas instaladas. Si no es así, puedes instalarlas usando `pip`:

bash

Copy code

```
pip install yfinance pandas scikit-learn
```

```
# Agregar columnas de características basadas en la fecha
df['Year'] = df['Date'].dt.year
df['Month'] = df['Date'].dt.month
df['Day'] = df['Date'].dt.day
df['Weekday'] = df['Date'].dt.weekday

# Preparar datos para el modelo
X = df[['Year', 'Month', 'Day', 'Weekday']].values
y = df['Price'].values

# Crear y entrenar modelo de regresión lineal
model = LinearRegression()
model.fit(X, y)

# Obtener la última fecha en el DataFrame
last_date = df['Date'].iloc[-1]

# Generar fechas para las próximas 4 semanas
forecast_dates = [last_date + timedelta(weeks=i) for i in range(1, 5)]

# Crear datos de características para las fechas de pronóstico
forecast_features = np.array([[date.year, date.month, date.day, date.weekday
```

Aquí tienes un ejemplo de cómo podrías hacer esto:

python

Copy code

```
import yfinance as yf
import pandas as pd
from datetime import datetime, timedelta
from sklearn.linear_model import LinearRegression
import numpy as np

# Definir el ticker de Bancolombia en Yahoo Finance
ticker = "CIB"

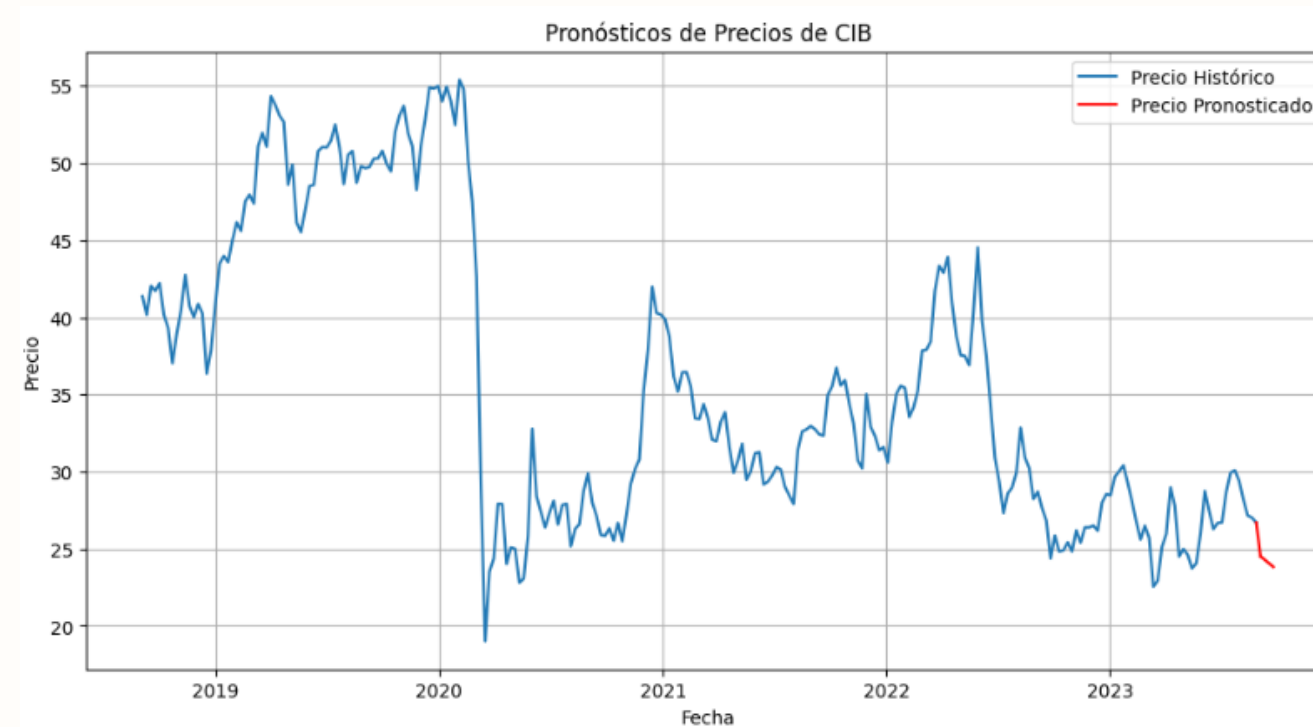
# Definir fechas de inicio y fin
end_date = datetime.today().date()
start_date = end_date - timedelta(days=5*365)

# Descargar datos desde Yahoo Finance
data = yf.download(ticker, start=start_date, end=end_date, interval="1wk")

# Seleccionar columna "Close" (precio de cierre)
data = data["Close"]

# Crear un DataFrame con fechas y precios
df = pd.DataFrame({'Date': data.index, 'Price': data.values})
```

Validación de los resultados



De acuerdo con el gráfico, se observa una alta volatilidad en el precio promedio semanal de Bancolombia, esto es de esperar sabiendo que se trata de una acción transada en bolsa. Se destaca que el pronóstico obtenido del modelo indica que la tendencia a la baja se mantendrá durante las siguientes 4 semanas, llegando a un mínimo de 23.82USD por acción.

Validación de los resultados

```
=====
                        OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.441
Model:                  OLS    Adj. R-squared:      0.434
Method:                 Least Squares    F-statistic:      67.47
Date:                   Fri, 01 Sep 2023    Prob (F-statistic):  3.29e-32
Time:                   15:56:15    Log-Likelihood:     -880.43
No. Observations:       261    AIC:              1769.
Df Residuals:           257    BIC:              1783.
Df Model:                3
Covariance Type:        nonrobust
=====
                        coef    std err          t      P>|t|      [0.025      0.975]
-----
const          8814.9706    619.387     14.232     0.000     7595.250     1e+04
x1              -4.3425      0.306    -14.172     0.000      -4.946     -3.739
x2              -0.6059      0.133     -4.552     0.000      -0.868     -0.344
x3              -0.0328      0.050     -0.656     0.513      -0.131      0.066
x4               0           0         nan         nan         0         0
=====
Omnibus:            1.335    Durbin-Watson:      0.089
Prob(Omnibus):      0.513    Jarque-Bera (JB):    1.381
Skew:               -0.170    Prob(JB):            0.501
Kurtosis:           2.891    Cond. No.            inf
=====
```

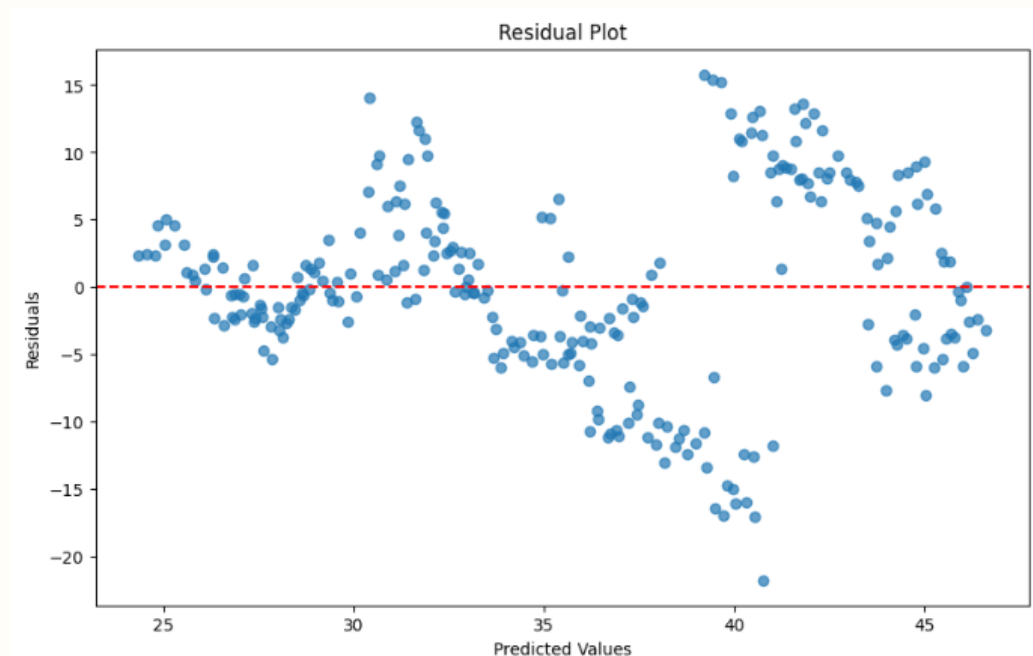
Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 0. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

- Se observa que el coeficiente de determinación del modelo es de 0.441, lo cual indica que las variables independientes explican en un 44,1% el comportamiento del precio promedio semanal de Bancolombia.
- Las variables Day y Weekday no son significativas, es decir, no tienen un efecto sobre el precio de la acción.
- No obstante, hay que resaltar que las variables explicativas (Year, Month, Day, Weekday) podrían estar aportando la misma información al modelo, generando problemas de multicolinealidad.
- El estadístico Durbin-Watson es menor a 2, indicando una autocorrelación serial positiva, lo cual indica cierta dependencia entre los errores.

Validación de los resultados

Distribución del error



La gráfica de distribución de los residuos resalta una amplia varianza en los residuos, indicando posibles problemas de heterocedasticidad a pesar de que el valor promedio tiende a cero.

Valor promedio del error

```
1 print(round(np.mean(residuals), 2))
```

-0.0

Se observa que el valor promedio del error tiende a cero, lo cual es acorde con los supuestos de los modelos de MCO.

Validación de los resultados

Test de White

```
LM Estadístico: 59.137739792585656
P-valor del LM: 1.9648641225940722e-09
F Estadístico: 8.170352657900473
P-valor del F: 1.2237426432618652e-10
Heterocedasticidad detectada: Rechaza la hipótesis nula
```

Los resultados del test de White indican que el error presenta problemas de heterocedasticidad, lo cual indica que este no sigue una distribución normal (media cero y varianza constante).

Test de VIF

	Variable	VIF
0	X1	7.658239
1	X2	4.599299
2	X3	4.207281
3	X4	NaN

Se observa que los resultados del estadístico VIF para las variables independientes del modelo son mayores a 1, lo cual indica que evidentemente el modelo presenta problemas de multicolinealidad.

Conclusiones

- Si bien el modelo es capaz de proporcionar los pronósticos del precio promedio semanal para Bancolombia, los análisis realizados indican que estos pronósticos no son fiables, pues en primer lugar la selección de variables es incorrecta debido a la alta multicolinealidad detectada entre ellas. A partir de allí se detecta una serie de problemas del modelo asociados a la autocorrelación de los errores y heterocedasticidad. El resultado final es un modelo con una significancia global bastante baja (44,1%).
- Además, cabe mencionar que la metodología utilizada tampoco es la adecuada, pues las series de tiempo asociadas a las acciones y demás instrumentos del mercado de valores son denominadas como "series explosivas", es decir, que presentan una alta varianza. En consecuencia, la implementación de un modelo ARCH o GARCH, los cuales están diseñados para el pronóstico de series de este tipo, puede llegar a ser más eficiente.
- Por último, se destaca que el modelo implementado no fue pre-entrenado, lo cual es una buena práctica en la implementación de modelos con el fin de evitar el sobreajuste.