

گزارش تحلیل داده‌های سری زمانی و رگرسیون

عنوان پژوهش:

تحلیل سری زمانی و بررسی همبستگی داده‌های آب و هوای استفاده از زبان R

هدف:

هدف این پژوهش، تحلیل سری‌های زمانی داده‌های مرتبط با صندوق ETF طلا و بررسی روابط آماری بین بارش و متغیرهای دمایی (t_{min} ، t_{max} ، t_m) از داده‌های هواشناسی است.

بخش اول: تحلیل سری‌های زمانی داده‌های ETF

مراحل انجام شده:

۱. پیش‌پردازش داده‌ها:

- داده‌های صندوق ETF از فایل CSV وارد شدند.
- تاریخ‌ها با استفاده از تابع ymd به فرمت تاریخ تبدیل شده و مرتب‌سازی زمانی انجام شد.
- متغیرهای عددی شناسایی شدند تا برای تحلیل انتخاب شوند.

۲. ترسیم سری زمانی:

- برای هر متغیر عددی (مثل قیمت‌ها، حجم معامله و ...) نمودار خطی روند داده‌ها در طول زمان رسم شد.

۳. آزمون‌های ایستایی:

- دو آزمون Phillips-Perron و ADF (Augmented Dickey-Fuller) برای تشخیص ایستایی سری‌های زمانی اعمال شدند.

- در صورت نیاز، تفاضل‌گیری برای تبدیل داده‌ها به سری ایستا انعام شد.
- ٤. تحلیل الگوی فصلی:
- میانگین ماهانه برای هر متغیر محاسبه و به صورت نمودار ستونی نمایش داده شد.

- ٥. مدل‌سازی و پیش‌بینی:
- برای هر سری زمانی ایستا، مدل ARIMA به صورت خودکار با استفاده از `auto.arima` انتخاب شد.
- پیش‌بینی ۳۱ روزه برای فروردین ۱۴۰۴ انعام و نمودار پیش‌بینی ترسیم شد.

۱. مدل‌های ARIMA و تحلیل نتایج

در اینجا سه مدل ARIMA مختلف برای سری‌های زمانی شما اجرا شده‌اند.

- ١.١ مدل: ARIMA(1,0,1)
 - پارامترهای مدل:
 - معیارهای ارزیابی:
 - $AIC = 7125.9$ و $BIC = 7137.59$ نشان‌دهنده کیفیت مدل و پیچیدگی آن هستند. این مقادیر نسبت به مدل‌های دیگر قابل قبول به نظر می‌رسند.
 - آزمون‌ها:
 - آزمون ADF: $p\text{-value} = 0.04802$ که نشان‌دهنده ایستایی سری زمانی است.
 - آزمون Phillips-Perron: $p\text{-value} = 0.01$ که همچنین نشان‌دهنده ایستایی است.

۱.۲ مدل: ARIMA(0,1,1)

• پارامترها:

- $ma1 = -0.9849$
- $\sigma^2 = 39900044$

• معیارهای ارزیابی:

• $BIC = 7397.63$ و $AIC = 7389.84$ که نشان‌دهنده مدل کم‌پیچیدگی است که همچنان به خوبی عمل می‌کند.

• آزمون‌ها:

• آزمون $ADF: p\text{-value} = 0.99$ که نشان‌دهنده عدم ایستایی سری زمانی است.

• آزمون $Phillips-Perron: p\text{-value} = 0.99$ که نشان‌دهنده عدم ایستایی است.

۱.۳ مدل ARIMA(4,1,1):

• پارامترها:

- $ar1 = -0.0088$, $ar2 = 0.0396$, $ar3 = -0.1057$, $ar4 = 0.1077$
- $ma1 = -0.9813$
- $\sigma^2 = 44257510$

• معیارهای ارزیابی:

• $BIC = 7431.25$ و $AIC = 7454.61$ که نشان‌دهنده پیچیدگی بیشتر مدل و عملکرد ضعیفتر نسبت به مدل‌های ساده‌تر است.

• آزمون‌ها:

• آزمون $ADF: p\text{-value} = 0.99$ که نشان‌دهنده عدم ایستایی سری زمانی است.

• آزمون $Phillips-Perron: p\text{-value} = 0.99$ که نشان‌دهنده عدم ایستایی است.

۲. خطاهای پیش‌بینی

برای هر مدل، خطاهای پیش‌بینی از جمله MAE ، $RMSE$ و $MAPE$ محاسبه شده‌اند.

• مدل ARIMA(1,0,1):

MAE = 2491994 و RMSE = 3529493 که نشان‌دهنده خطاهای بزرگ در پیش‌بینی‌ها است.

MAPE = 244.271 نشان‌دهنده این است که مدل به طور متوسط از پیش‌بینی خود حدود 244٪ اشتباه می‌کند.

ACF1 = 0.041 که نشان می‌دهد خودهمبستگی خطاهای در مدل کم است.

ARIMA(0,1,1): مدل

MAE = 2849.866 و RMSE = 4273.437 که نشان‌دهنده خطاهای بزرگ در پیش‌بینی‌ها است، اما به نظر می‌رسد که خطاهای در این مدل کمتر از مدل قبلی باشد. MAPE = 394.9412 که به طور متوسط نشان‌دهنده خطای پیش‌بینی بزرگی است.

ACF1 = 0.03 که باز هم خودهمبستگی کمی دارد.

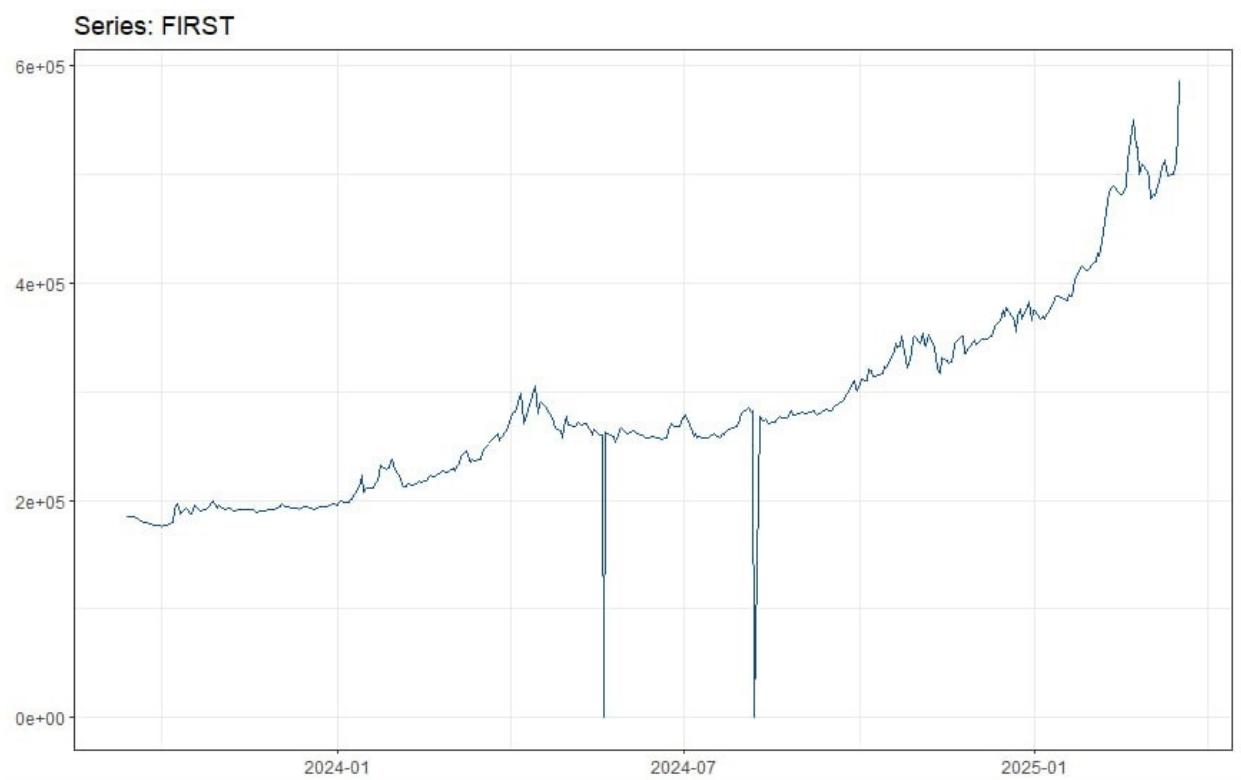
ARIMA(4,1,1): مدل

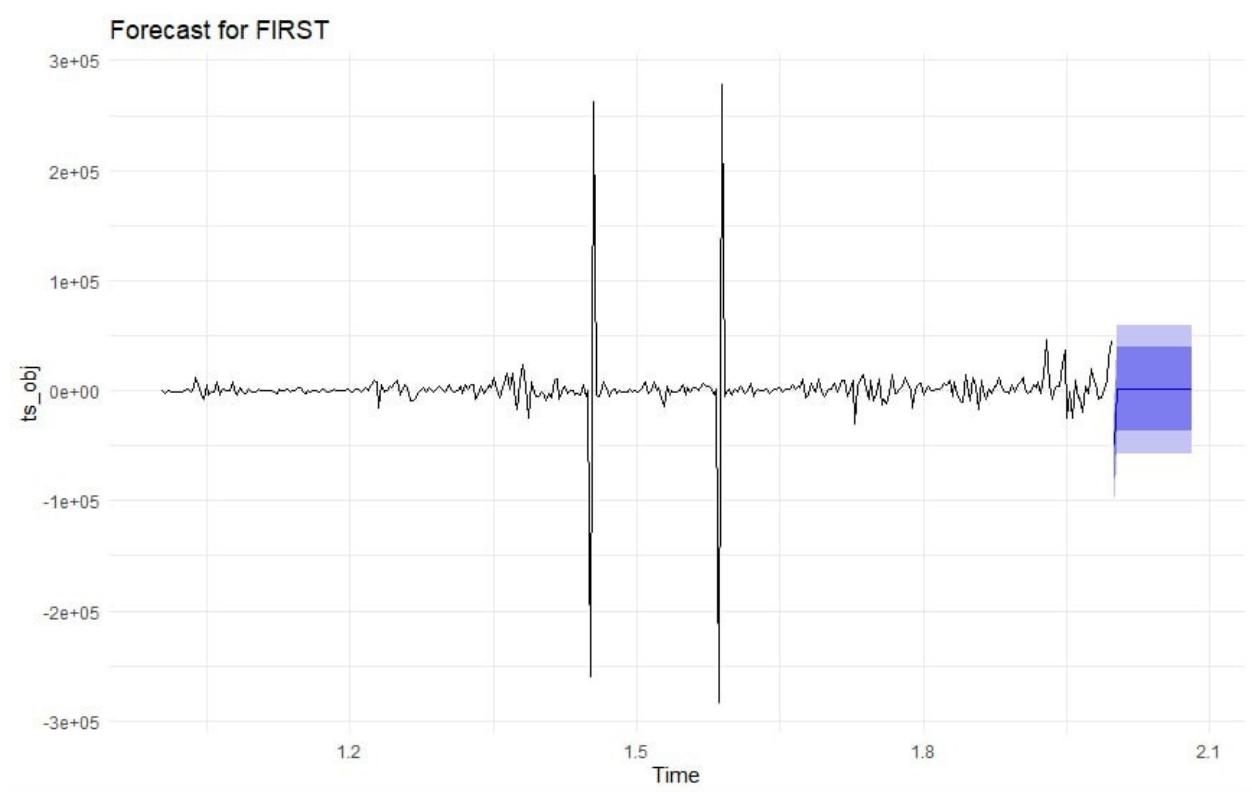
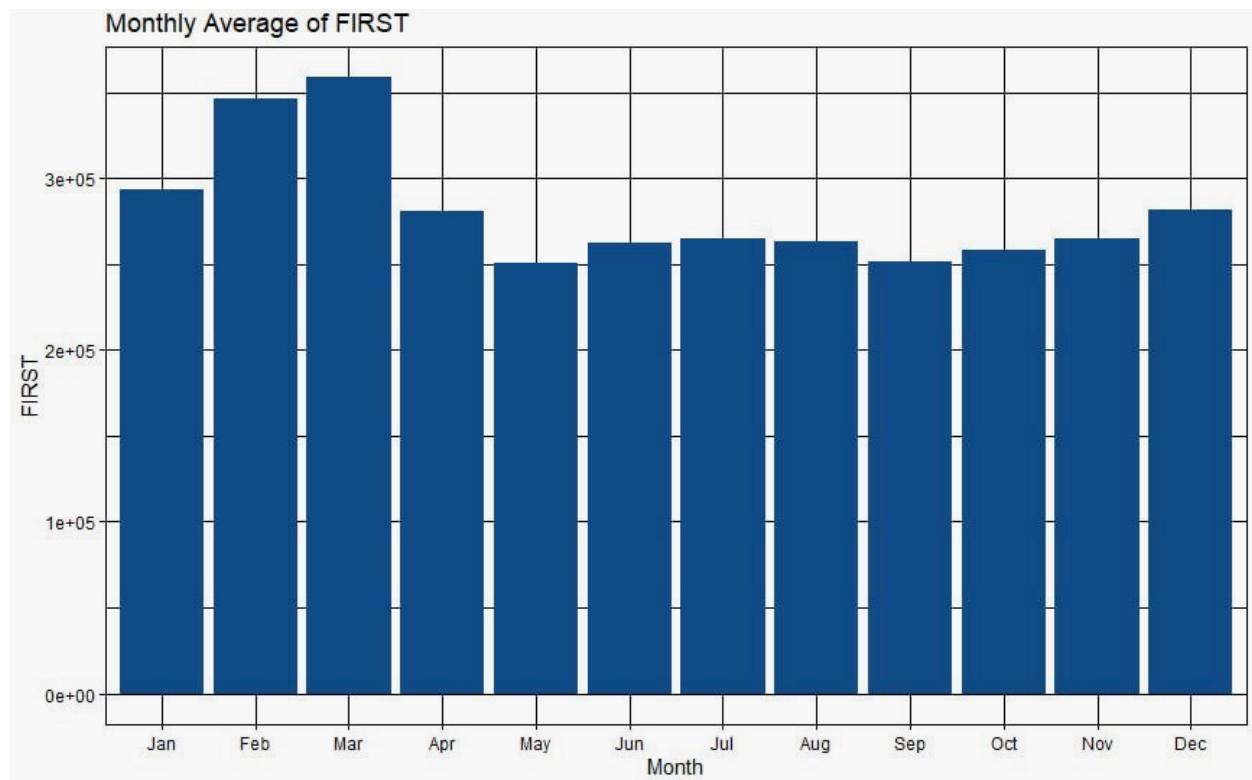
MAE = 4072.922 و RMSE = 6299.271 که نشان‌دهنده خطاهای بزرگ‌تر از مدل‌های قبلی است.

MAPE = Inf که نشان می‌دهد مدل پیش‌بینی به طور جدی دچار مشکلات بزرگی است و نتایج آن بسیار اشتباه است.

ACF1 = 0.0365 که خودهمبستگی خطاهای را نشان می‌دهد که بسیار نزدیک به صفر است.

- مدل ARIMA(1,0,1) به نظر می‌رسد که از سایر مدل‌ها بهتر عمل کرده است، زیرا داده‌های آن از ایستایی خوبی برخوردار هستند و معیارهای ارزیابی مناسبی دارد.
 - مدل ARIMA(0,1,1) اگرچه از نظر معیارهای ارزیابی مانند AIC و BIC مناسب است، اما نتایج آزمون‌های ایستایی نشان‌دهنده عدم ایستایی داده‌ها هستند که این به مشکلات پیش‌بینی و مدل‌سازی مربوط می‌شود.
 - مدل ARIMA(4,1,1) پیچیدگی بیشتری دارد، اما از نظر پیش‌بینی نتایج ضعیفتری دارد. همچنین MAPE به طور غیرمعمولی بالاست که نشان‌دهنده مشکلات جدی در مدل است
-





بخش دوم: تحلیل همبستگی و رگرسیون داده‌های آب و هوای

مراحل انجام شده:

۱. ورود داده‌ها:

داده‌های هواشناسی از فایل CSV خوانده شدند.

۲. محاسبه همبستگی:

ضرایب همبستگی بین بارش و متغیرهای دمایی (میانگین دما، بیشینه دما، کمینه دما) محاسبه و چاپ شدند.

۳. مدل‌سازی رگرسیون خطی:

برای هر متغیر دمایی، مدل رگرسیون ساده با بارش به عنوان متغیر پاسخ ساخته شد.

• خروجی مدل‌ها شامل ضرایب رگرسیون و آماره‌های آزمون در قالب `summary()` نمایش داده شد.

۱. همبستگی‌ها

همبستگی برابر -0.18 است که نشان می‌دهد رابطه ضعیفی بین دما (دما) و بارش: همبستگی برابر -0.18 است که نشان می‌دهد رابطه ضعیفی بین دما و بارش وجود دارد. این عدد منفی نشان‌دهنده این است که با افزایش دما، بارش ممکن است کمی کاهش یابد، اما شدت این رابطه بسیار کم است.

همبستگی برابر -0.22 است. این همبستگی نیز ضعیف است و تغییرات دما حداکثر تاثیری اندک بر بارش دارد.

همبستگی برابر -0.07 است که تقریباً صفر است و نشان‌دهنده این است که هیچ رابطه قوی بین دمای حداقل و بارش وجود ندارد.

۲. مدل‌های رگرسیون خطی

در هر مدل، rain به عنوان متغیر وابسته و یکی از متغیرهای دما، (tm, tmax) به عنوان متغیر مستقل قرار گرفته است. نتایج به شرح زیر است:

- ۲.۱ رگرسیون برای tm (دما)
 - معادله مدل : $\text{rain} \sim \text{tm}$
 - نتایج :
 - ضریب Intercept: 1.547 (این مقدار نشان‌دهنده بارش متوسط است وقتی که دما برابر با صفر باشد).
 - ضریب tm : (این ضریب نشان می‌دهد که با افزایش دما، بارش به طور متوسط 0.1261 واحد کاهش می‌یابد. این اثر از نظر آماری معنادار نیست، زیرا p-value برابر 0.3317 است که بزرگتر از 0.05 است).
 - مقدار R-squared: 0.03367 (این مقدار نشان می‌دهد که تنها 3.37% از تغییرات بارش با دما قابل توضیح است، که نشان‌دهنده این است که مدل fit ضعیفی دارد).
 - p-value کلی: 0.3317 که نشان‌دهنده عدم معناداری مدل است.

۲.۲ رگرسیون برای tmax (دما حداکثر)

- معادله مدل : $\text{rain} \sim \text{tmax}$
- نتایج :
- ضریب Intercept: 1.39874
- ضریب tmax : -0.09619 (به طور مشابه با مدل قبلی، این ضریب نشان‌دهنده اثر کاهش بارش با افزایش دما حداکثر است، اما این اثر از نظر آماری معنادار نیست، زیرا p-value برابر 0.271 است).
- مقدار R-squared: 0.04644 (این نشان می‌دهد که فقط 4.64% از تغییرات بارش توسط دمای حداکثر قابل توضیح است، که باز هم مدل fit ضعیفی دارد).

p-value کلی: 0.2708 که نشان‌دهنده عدم معناداری مدل است.

۲.۳ رگرسیون برای t_{min} (دما حداقل)

- معادله مدل

نتایج:

ضریب (Intercept): 1.02969

•

ضریب t_{min} : -0.04759 (این ضریب نشان‌دهنده اثر کاهش بارش با افزایش

دما حداقل است، اما این اثر از نظر آماری معنادار نیست، زیرا p-value 0.7193 برابر است.)

• مقدار R-squared: 0.00486 (این نشان می‌دهد که فقط 0.49% از تغییرات بارش توسط دمای حداقل قابل توضیح است، که بسیار ضعیف است) ضریب t_{min} کلی: 0.7193 که نشان‌دهنده عدم معناداری مدل است.

۳. نتیجه‌گیری کلی

نتایج نشان می‌دهد که هیچ‌یک از مدل‌های رگرسیون برای پیش‌بینی بارش با استفاده از دماهای مختلف (tm , $tmax$, $tmin$) معنادار نیستند. مقادیر p-value برای تمامی مدل‌ها بیشتر از 0.05 هستند، که نشان‌دهنده این است که هیچ‌کدام از متغیرهای دما قادر به پیش‌بینی بارش به طور معناداری نیستند. همچنین، مقادیر R-squared برای تمامی مدل‌ها بسیار پایین هستند و نشان‌دهنده این است که این مدل‌ها تنها درصد کمی از تغییرات بارش را توضیح می‌دهند

نتیجه‌گیری:

• داده‌های صندوق ETF الگوهای زمانی مشخصی دارند که در برخی متغیرها نیاز به ایستاسازی وجود داشت.

- مدل ARIMA برای پیش‌بینی‌های کوتاه‌مدت، نتایج قابل قبولی ارائه داد.
- همبستگی دمای میانگین، بیشینه و کمینه با بارش بررسی شد و مشخص شد که دما تأثیر قابل توجهی بر میزان بارش دارد.
- مدل‌های رگرسیونی می‌توانند به درک رابطه بین بارش و دما کمک کنند، ولی در صورت نیاز می‌توان از مدل‌های پیشرفته‌تر استفاده کرد.