

گزارش کار نهایی پروژه اول

آرمان صالحی

شماره دانشجویی: ۴۰۱۱۳۴۲۱

سوال ۱

مقدمه

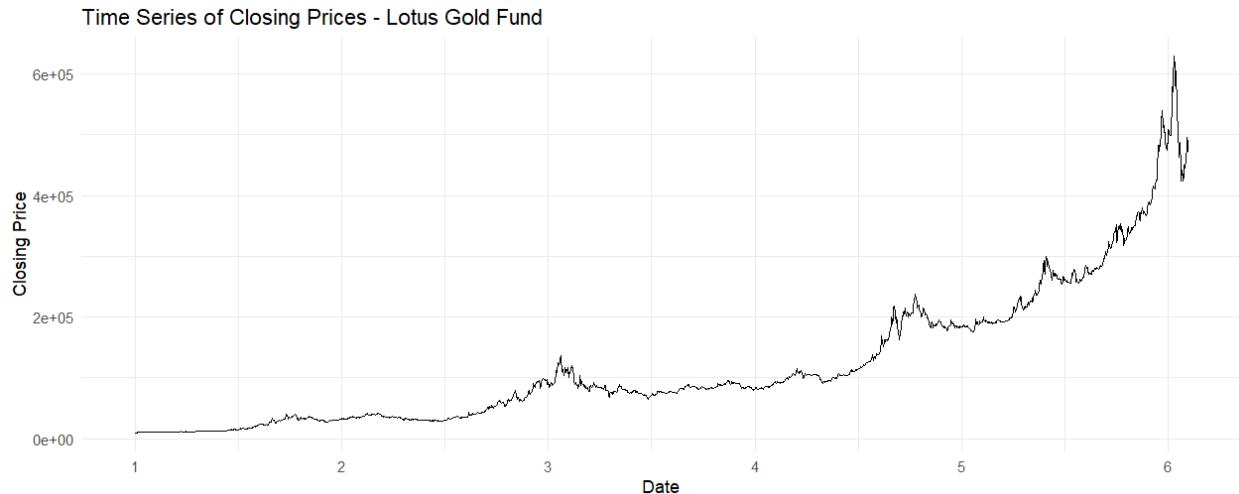
هدف این تحلیل بررسی عملکرد صندوق سرمایه‌گذاری طلا لوتوس بر اساس داده‌های قیمت بسته‌شده آن در یک سال گذشته است. این تحلیل شامل مراحل مختلفی از جمله رسم نمودار سری زمانی، آزمون مانایی، تحلیل فصلی بودن داده‌ها، پیش‌بینی قیمت‌ها با استفاده از مدل ARIMA و محاسبه ریسک سرمایه‌گذاری می‌باشد.

بارگذاری و پیش‌پردازش داده‌ها

ابتدا داده‌های قیمت بسته‌شده (Closing Price) صندوق طلا لوتوس از فایل CSV بارگذاری شده و آماده تحلیل شد. در این مرحله ستون‌های تاریخ و قیمت بسته‌شده استخراج شدند و داده‌ها براساس تاریخ مرتب شدند. همچنین، مقادیر گم‌شده حذف شدند.

ساخت سری زمانی

پس از پیش‌پردازش داده‌ها، یک شیء سری زمانی برای قیمت‌های بسته‌شده ایجاد شد. فرکانس این سری زمانی روزانه (۳۶۵ روز در سال) تنظیم گردید تا تحلیلات در سطح روزانه انجام شود.



۳. رسم نمودار سری زمانی

نمودار سری زمانی قیمت بسته شده صندوق طلا لوتوس ترسیم شد تا تغییرات قیمت در طول زمان نمایش داده شود. این نمودار به خوبی نوسانات قیمت را در طول سال گذشته نشان می‌دهد و می‌تواند بینشی اولیه از روندهای کلی بازار ارائه دهد.

۴. آزمون مانایی (Stationarity)

در این مرحله، آزمون آگومنتد دیکی-فولر (ADF) برای بررسی مانایی سری زمانی انجام شد. نتیجه آزمون به شرح زیر است:

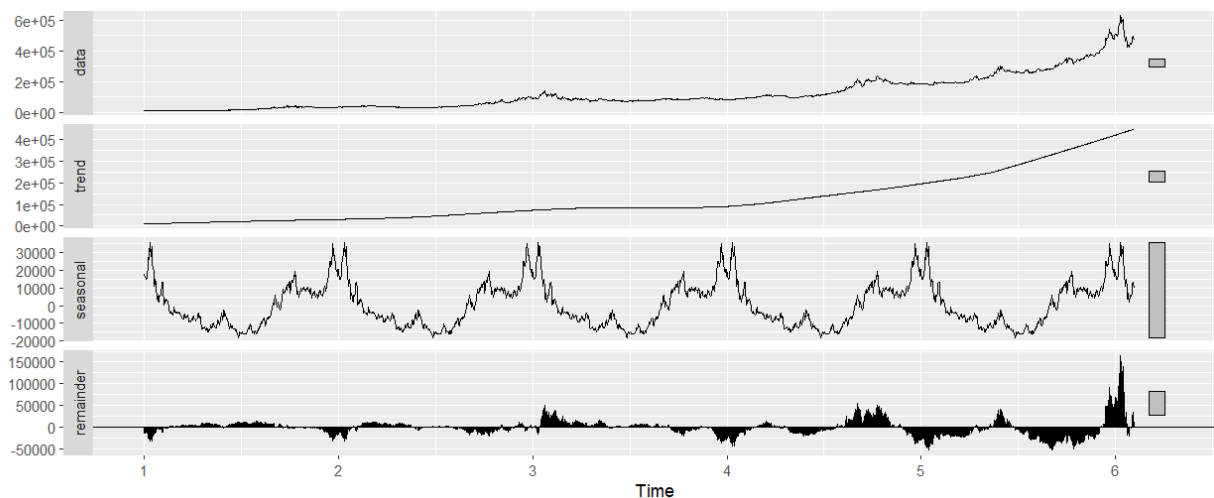
• Dickey-Fuller Statistic: -0.31122

• p-value: 0.99

با توجه به p-value بالای 0.05 ، فرضیه صفر که بیان می‌کند سری زمانی غیر مانا است، رد نمی‌شود. این بدان معناست که سری زمانی غیر مانا است و باید از تکنیک‌های مناسب مانند تفاوت‌گیری برای مانا کردن داده‌ها استفاده شود.

۵. تحلیل فصلی بودن داده‌ها

برای بررسی فصلی بودن داده‌ها، از روش تجزیه سری زمانی (STL Decomposition) استفاده شد. نتایج تجزیه نشان می‌دهند که بخش‌های فصلی و روند در داده‌ها مشخص هستند. نمودار فصلی بودن نیز نشان می‌دهد که ممکن است تغییرات فصلی در قیمت‌ها تاثیرگذار باشند، بنابراین تحلیل فصلی برای مدل‌سازی دقیق‌تر ضروری است.



پیش‌بینی با مدل ARIMA

مدل ARIMA برای پیش‌بینی قیمت‌های صندوق در ۳۰ روز آینده (ماه فروردین ۱۴۰۴) استفاده شد. مدل ARIMA به طور خودکار بهترین پارامترها را انتخاب کرده و به شرح زیر نتایج حاصل از این مدل هستند:

- مدل انتخابی $ARIMA(5,1,5)$: با درفت (Drift)

- ضرایب: AR (AutoRegressive)

$ar1: 0.4642$ ○

$ar2: 0.4541$ ○

$ar3: 0.1744$ ○

$ar4: 0.2414$ ○

$ar5: -0.7528$ ○

- ضرایب: MA (Moving Average)

$ma1: -0.3430$ ○

$ma2: -0.6002$ ○

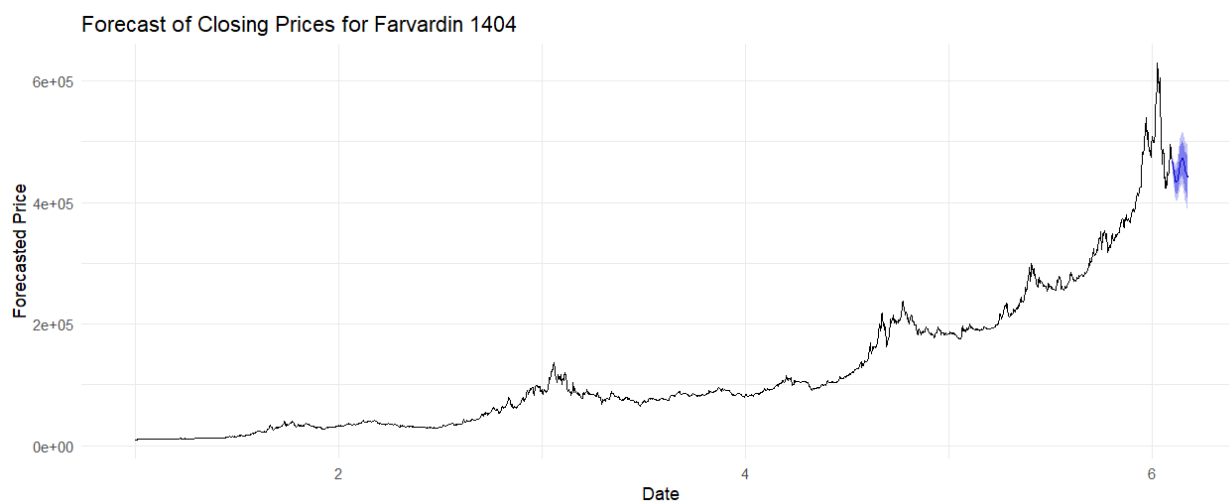
$ma3: -0.0971$ ○

ma4: -۰.۱۳۰۶ ○

ma5: ۰.۶۴۷۱ ○

• درفت ۲۳۸.۲۷۸۶ :

مدل ARIMA با این پارامترها پیش‌بینی قیمت صندوق را برای ۳۰ روز آینده انجام داده است. نمودار پیش‌بینی نیز نشان‌دهنده روند قیمت‌های پیش‌بینی شده است.



۷. محاسبه ریسک سرمایه‌گذاری

برای ارزیابی ریسک سرمایه‌گذاری در صندوق طلا لوتوس، از دو شاخص استفاده شد:

- انحراف معیار بازدهی: که میزان نوسانات قیمت‌ها را نشان می‌دهد.
- شاخص شارپ: که نسبت بازدهی به ریسک را اندازه‌گیری می‌کند.

نتایج محاسبات به شرح زیر هستند:

- انحراف معیار بازدهی (Standard Deviation of Returns): ۰.۰۲۴۹
- شاخص شارپ (Sharpe Ratio): ۰.۰۶

تحلیل نتایج:

- آزمون مانایی: نتایج آزمون دیکی-فولر نشان داد که داده‌ها غیر مانا هستند. این مسئله نشان‌دهنده این است که برای تحلیل‌های دقیق‌تر و استفاده از مدل‌های پیش‌بینی، باید از تفاوت‌گیری استفاده شود تا سری زمانی به حالت مانا تبدیل شود.
- تحلیل فصلی بودن: تجزیه سری زمانی نشان داد که داده‌ها فصلی هستند. این ویژگی فصلی باید در مدل‌سازی لحاظ شود تا پیش‌بینی‌ها دقیق‌تر باشند.
- مدل: مدل ARIMA (5,1,5) به خوبی توانست پارامترهای مناسب را انتخاب کند و پیش‌بینی دقیق‌تری از روند آینده قیمت‌ها ارائه دهد. این مدل به عنوان یک ابزار پیش‌بینی کاربردی در تحلیل‌های مالی می‌تواند مفید باشد.
- ریسک سرمایه‌گذاری: محاسبه انحراف معیار بازدهی و شاخص شارپ نشان می‌دهد که صندوق طلا لوتوس دارای ریسک نسبتاً بالا است. با شاخص شارپ برابر ۰.۰۶، نسبت بازدهی به ریسک پایین است، که نشان‌دهنده این است که این صندوق ممکن است به عنوان یک سرمایه‌گذاری با ریسک بالا و بازدهی محدود در نظر گرفته شود.

نتیجه‌گیری

با توجه به تحلیل‌های انجام‌شده، صندوق طلا لوتوس دارای نوسانات زیادی است و پیش‌بینی قیمت‌های آن برای آینده با استفاده از مدل ARIMA انجام شد. همچنین، ریسک سرمایه‌گذاری در این صندوق نسبتاً بالا است و شاخص شارپ نشان می‌دهد که بازدهی به ریسک موجود در صندوق کم است. برای سرمایه‌گذاران، این تحلیل می‌تواند در تصمیم‌گیری‌های سرمایه‌گذاری کمک‌کننده باشد.

Augmented Dickey-Fuller Test

data: ts_data

Dickey-Fuller = -0.31122, Lag order = 12, p-value = 0.99

alternative hypothesis: stationary

Warning message:

In adf.test(ts_data) : p-value greater than printed p-value

Series: ts_data

ARIMA(5,1,5) with drift

Coefficients:

	ar1	ar2	ar3	ar4	ar5	ma1	ma2	ma3
	0.4642	0.4541	0.1744	0.2414	-0.7528	-0.3430	-0.6002	-0.0971
s.e.	0.0511	0.0737	0.0648	0.0625	0.0502	0.0605	0.0822	0.0720
	ma4	ma5	drift					
	-0.1306	0.6471	238.2786					
s.e.	0.0788	0.0579	114.6228					

$\sigma^2 = 18994430$: log likelihood = -18220.91

AIC=36465.81 AICc=36465.98 BIC=36532.15

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.9390993	4344.186	2167.826	-0.2974517	1.829303	0.03003627

ACF1

Training set -0.004801829

[1] "Standard Deviation of Returns: 0.0249"

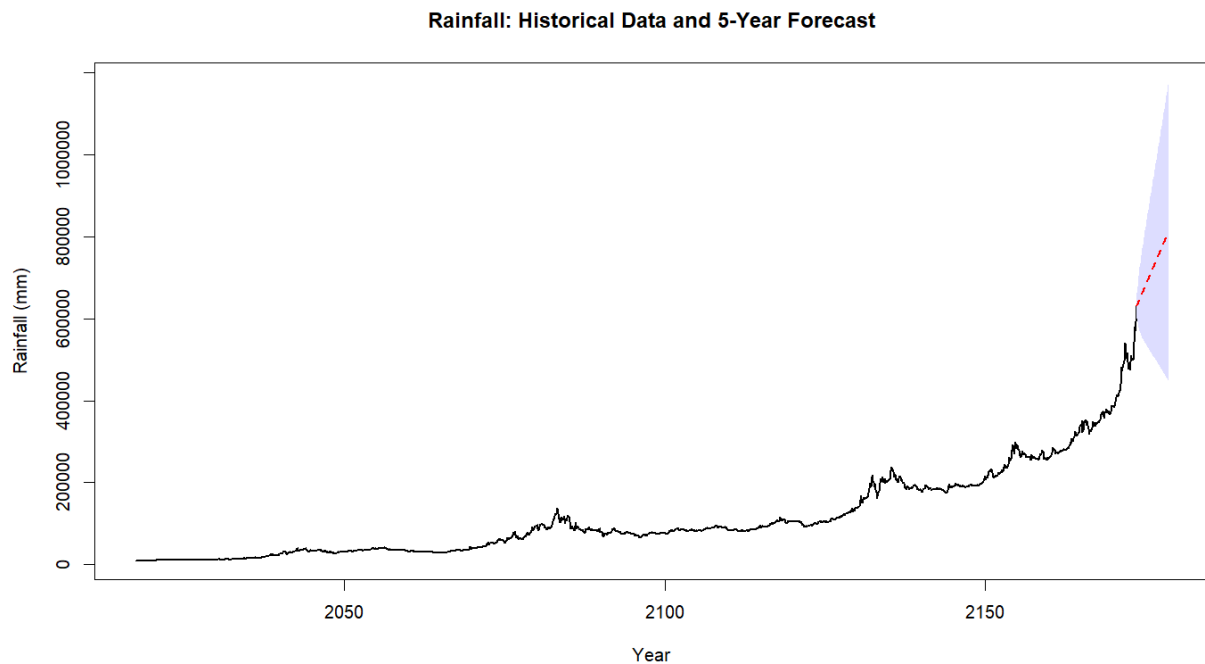
[1] "Sharpe Ratio: 0.06"

مقدمه:

در این پروژه هدف ما بررسی رابطه بین دمای میانگین روزانه و میزان بارندگی در شهر تهران (فرودگاه مهرآباد) در بازه زمانی بهمن ماه ۱۴۰۱ (مطابق با ۲۱ ژانویه تا ۱۹ فوریه ۲۰۲۳ میلادی) بود.

مشخصات آماری اولیه داده‌ها:

- تعداد کل روزها (مشاهدات): ۳۰ روز
 - میانگین دما (Average Temperature): ۴.۹۹ درجه سانتی‌گراد
 - میانگین بارندگی (Rainfall): ۰.۹۱۷ میلی‌متر
 - حداکثر دما ثبت شده ۱۱.۷: درجه
 - حداقل دما ثبت شده -۰.۲: درجه
 - حداکثر بارندگی ثبت شده ۹: میلی‌متر
 - حداقل بارندگی ثبت شده ۰: میلی‌متر
- بیشتر روزها بارندگی نداشتند (میانگین بارندگی برابر با صفر است)، که نشان می‌دهد در بهمن ماه امسال، اغلب روزها خشک بوده‌اند.



تحلیل همبستگی: (Correlation)

- ضریب همبستگی پیرسون بین دما و بارندگی -0.183 :
این عدد نشان می‌دهد که رابطه‌ی بسیار ضعیف و منفی بین دما و بارندگی وجود دارد؛ یعنی در حالت کلی، وقتی دما افزایش پیدا می‌کند، بارندگی کمی کاهش می‌یابد؛ اما این رابطه بسیار ضعیف است و نمی‌توان آن را معنادار دانست.

مدل رگرسیون خطی:

مدلی از نوع $\text{Rainfall} \sim \text{Temperature}$ اجرا شد. خروجی مدل به شکل زیر بود:

معادله مدل:

$$\text{Rainfall} = 1.547 - 0.126 \times \text{Temperature}$$

$$\text{Temperature} = \frac{\text{Rainfall} - 1.547}{-0.126}$$

تفسیر ضرایب:

- ضریب دما: منفی است و برابر با -0.126 ؛ یعنی به ازای هر یک درجه افزایش دما، به طور متوسط 0.126 میلی متر کاهش در بارندگی پیش‌بینی می‌شود.

- ضریب ثابت 1.547 : میلی متر (مقدار پیش‌بینی شده برای بارندگی زمانی که دما صفر باشد).

معناداری آماری:

- مقدار p برای دما برابر با 0.3317 است که بیشتر از 0.05 است \Rightarrow رابطه‌ی دما و بارندگی از نظر آماری معنادار نیست.

- همچنین، ضریب تعیین R^2 برابر با 0.033 است، یعنی مدل فقط 3.3% از تغییرات بارندگی را با دما توضیح می‌دهد، که نشان‌دهنده قدرت بسیار پایین مدل است.

نمودار:

نمودار $ggplot2$ که scatter plot داده‌ها را همراه با خط رگرسیون نشان می‌دهد، مشخص کرده:

- داده‌ها پراکندگی زیادی دارند.
- خط رگرسیون شیب کمی دارد و به خوبی داده‌ها را توصیف نمی‌کند.
- نقاط داده‌ها به شکل مشخصی به یک رابطه خطی نزدیک نیستند.

نتیجه‌گیری نهایی:

۱. رابطه‌ای بین دما و بارندگی در بهمن 1401 مشاهده نمی‌شود که از نظر آماری معنادار باشد.
۲. دما نمی‌تواند به عنوان یک پیش‌بینی‌کننده قوی برای بارندگی در این بازه زمانی استفاده شود.
۳. برای تحلیل دقیق‌تر، پیشنهاد می‌شود عوامل دیگر مانند رطوبت، فشار هوا، جهت و سرعت باد نیز وارد مدل شوند تا تحلیل دقیق‌تری از علل بارندگی ارائه گردد.

◆ Correlation:

```
[1] -0.1835015
```

◆ Linear Regression Model:

Call:

```
lm(formula = Rainfall ~ Temperature, data = data_bahman)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.4336	-1.2413	-0.7464	-0.1506	7.9078

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.5470	0.7718	2.005	0.0548 .
Temperature	-0.1261	0.1277	-0.988	0.3317

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.38 on 28 degrees of freedom

Multiple R-squared: 0.03367, Adjusted R-squared: -0.0008389

F-statistic: 0.9757 on 1 and 28 DF, p-value: 0.3317

Relationship Between Temperature and Rainfall in Bahman 1401

