**FestivalSync**

**Monitoraggio e Forecasting**

|  |  |  |
| --- | --- | --- |
| Versione Documento |  | Data Documento |
| 1.0 |  | 05/12/2024 |

SOMMARIO

[1. STORIA DEL DOCUMENTO 2](#_Toc184321226)

[2. INTRODUZIONE 2](#_Toc184321227)

[3. Monitoraggio e Observability 3](#_Toc184321228)

[3.1 Prometheus 3](#_Toc184321229)

[3.2 Grafana 3](#_Toc184321230)

[3.3 cAdvisor 3](#_Toc184321231)

[4. Forecasting 3](#_Toc184321232)

[4.1 Risultati 4](#_Toc184321233)

[4.1.1 CPU 4](#_Toc184321234)

[4.1.2 Latenza 8](#_Toc184321235)

# STORIA DEL DOCUMENTO

Le informazioni contenute in questo documento aggiornano e sostituiscono quelle contenute nelle versioni precedenti. Ogni modifica del documento sarà controllata nella apposita procedura.

|  |  |  |  |
| --- | --- | --- | --- |
| **Versione** | **Data** | **Descrizione della Revisione** | **Modificato da** |
| 1.0 | 05/12/2024 | Prima versione | Carlo Blatti |
|  |  |  |  |

# INTRODUZIONE

Nel contesto del progetto FestivalSync, il monitoraggio e l'osservabilità dell'infrastruttura e delle applicazioni è di fondamentale importanza per garantire la stabilità del sistema e la visibilità degli eventi che si verificano nei microservizi. Per raggiungere questo scopo, sono stati utilizzati Prometheus, Grafana e cAdvisor per raccogliere metriche, analizzarle e visualizzarle. Vediamo come questi strumenti sono stati integrati e configurati.

# Monitoraggio e Observability

## 3.1 Prometheus

Prometheus è stato scelto come sistema di monitoraggio principale per raccogliere metriche dai vari microservizi del progetto. Viene utilizzato per monitorare la salute delle applicazioni, le prestazioni e l'utilizzo delle risorse di sistema.

**Configurazione**:

* Micrometer e Spring Boot: I microservizi FestivalSync, come artist-service, event-service, e ticket-service, utilizzano Micrometer per esporre metriche in un formato compatibile con Prometheus. Micrometer è integrato con Spring Boot per facilitare la raccolta di metriche e il collegamento a Prometheus. La dipendenza è configurata nei file pom.xml.

**Endpoint per le Metriche:**

* Micrometer configura un endpoint /actuator/prometheus per ogni microservizio. Prometheus scrappa questi endpoint a intervalli regolari per raccogliere le metriche esposte. Ogni servizio, quindi, espone le metriche relative al carico CPU, memoria, numero di richieste, latenza, ecc.

## Grafana

Grafana viene utilizzato per visualizzare le metriche raccolte da Prometheus in modo grafico e intuitivo. L'integrazione di Grafana con Prometheus offre un'interfaccia per costruire dashboard personalizzate che mostrano l'andamento delle metriche chiave.

**Benefici di Grafana**:

* **Identificazione delle Anomalie**: Le visualizzazioni di Grafana aiutano a identificare trend sospetti, come un picco nelle latenze o un aumento degli errori delle API.
* **Alert Grafici**: Grafana può anche essere configurato per generare allarmi basati sulle metriche raccolte, con notifiche in tempo reale.

## cAdvisor

**cAdvisor (Container Advisor)** è uno strumento di monitoraggio specifico per i container Docker. cAdvisor è utilizzato per raccogliere metriche a livello di container Docker, fornendo visibilità sul consumo delle risorse da parte dei container dei vari microservizi.

**Configurazione**:

* **Integrazione con Docker**: cAdvisor viene eseguito come container Docker e monitora automaticamente tutti gli altri container in esecuzione sullo stesso host. Le metriche raccolte da cAdvisor vengono inviate a Prometheus.

# Forecasting

Per le metriche prese in esame, la latenza relativa alle chiamate HTTP/REST e all’utilizzo della CPU, è stato implementato un predittore **ARIMA**.

Attraverso degli script python è stato interrogato prometheus per ottenere i valori di interesse e questi sono stati salvati in un file csv dopo aver fatto pulizia dei dati non consistenti come i valori NaN o +Inf.

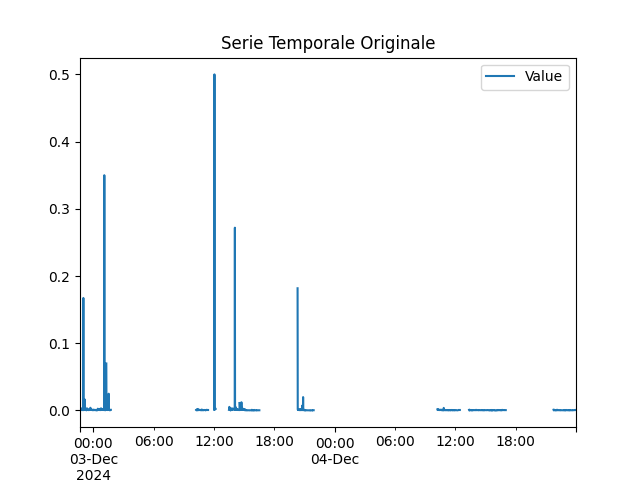
Poi sono stati creati due script python, uno per ogni metrica in esame, che implementano dei predittori ARIMA.

## 4.1 Risultati

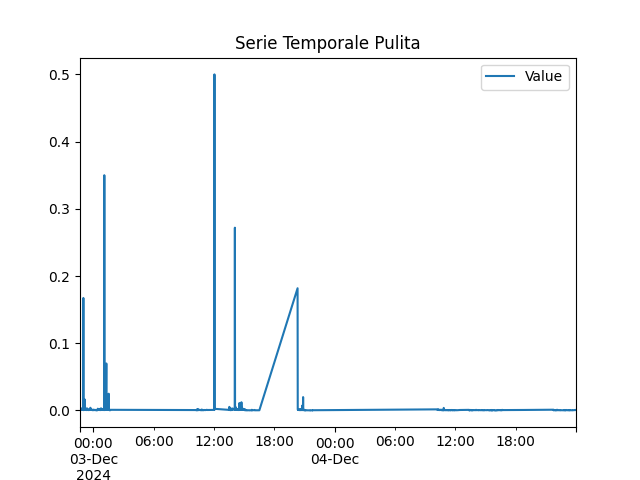
Di seguito i risultati ottenuti dal predittore.

### 4.1.1 CPU

Per quel che riguarda la CPU è stato preso in esame un campione di dati raccolti tra il 02/12/2024 e lo 04/12/2024.



Data l’inconsistenza dei dati è stata applicata interpolazione alla serie originale, il risultato è stato il seguente:

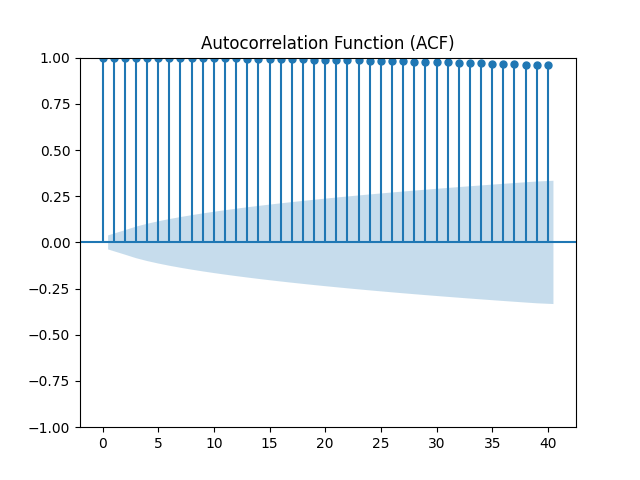


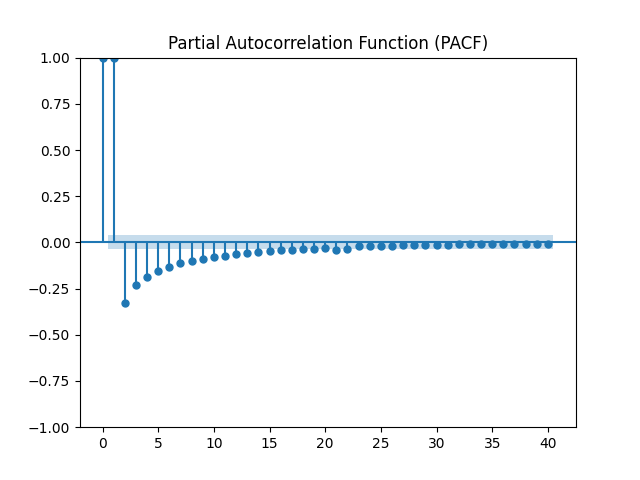
#### 4.1.1.2 Test di Dickey-Fuller (ADF)

Il **Test di Dickey-Fuller** è utilizzato per valutare se la serie temporale è stazionaria, cioè se ha proprietà statistiche costanti nel tempo. I risultati del test ADF sono:

* **Statistic ADF**: -3.122171
* **P-value**: 0.024955
* **Valori critici (1%, 5%, 10%)**: rispettivamente -3.432847, -2.862643, e -2.567357

Il **valore p** è **0.024955**, che è inferiore a 0.05, il che suggerisce che possiamo rifiutare l'ipotesi nulla di non stazionarietà con un livello di significatività del 5%. Inoltre, il valore **ADF statistic (-3.122171)** è inferiore al valore critico al livello del 5% (-2.862643), confermando ulteriormente che la serie è **stazionaria**. Pertanto, la serie è adatta per la modellazione ARIMA senza bisogno di ulteriori differenziazioni.





#### 4.1.1.3 Sommario del Modello

**p** (numero di termini autoregressivi): 2  
La grafica PACF mostra il primo punto significativo vicino a lag 2.

**d** (grado di differenziazione): 0  
Il test ADF ha confermato che la serie è già stazionaria, quindi non è necessaria differenziazione.

**q** (numero di termini media mobile): 1  
La grafica ACF mostra il primo punto significativo vicino a lag 1.

Il modello ARIMA scelto per questa serie quindi è **ARIMA(2,0,1)**, con i seguenti dettagli:

* **Numero di Osservazioni**: 2120
* **Log Likelihood**: 18181.273
* **AIC**: -36352.547 (il valore più basso di AIC suggerisce un buon adattamento del modello)
* **BIC**: -36324.251
* **HQIC**: -36342.188

**Coefficiente di Regressione**

* **Constante (const)**: 0.0112, con **p-value** molto piccolo (0.000), indicando che il termine è significativo.
* **AR Coefficienti**:
  + **ar.L1**: 1.9535, positivo e significativo (p-value ≈ 0.000)
  + **ar.L2**: -0.9536, negativo e significativo (p-value ≈ 0.000)
* **MA Coefficiente**:
  + **ma.L1**: 0.0648, anch'esso significativo con un p-value molto piccolo.

**Qualità dell'Adattamento**

* **Sigma²** (Varianza residua): 2.063e-09, indicando una varianza molto piccola nei residui, suggerendo che il modello è riuscito a catturare gran parte della variabilità della serie.
* **Ljung-Box Test (L1)** (Q): 1.71 con **Prob(Q)** di 0.19, suggerendo che i residui non sono correlati e il modello ARIMA ha catturato efficacemente la correlazione nella serie.
* **Jarque-Bera (JB)**: 3650980.13 con **Prob(JB)** di 0.00, suggerendo che i residui non seguono una distribuzione normale (cosa non insolita per dati finanziari o serie molto rumorose).
* **Heteroskedasticity (H)**: Valore di 0.00 suggerisce una probabile eteroschedasticità, ossia la varianza dei residui non è costante nel tempo.

**Metriche di Errore**

* **Mean Squared Error (MSE)**: 9.3898e-06
* **Root Mean Squared Error (RMSE)**: 0.00306

Questi valori suggeriscono un **errore medio quadratico basso**, il che implica che il modello si adatta bene ai dati storici e produce stime precise.

**Statistiche della Serie**

* **Media** della serie: 0.009056
* **Deviazione standard**: 0.018762
* **Minimo e Massimo**: Rispettivamente 0.000520 e 0.067515

Questi valori ci forniscono una panoramica della distribuzione dei dati, suggerendo una serie con **bassa varianza**, e un valore massimo abbastanza superiore rispetto alla media, suggerendo possibili outlier.

#### 4.1.1.4 Conclusioni Finali

1. **Stazionarietà**: La serie temporale è risultata stazionaria sulla base del test ADF. Ciò rende appropriato l'uso di un modello ARIMA senza necessità di ulteriori differenziazioni.
2. **Adattamento del Modello**: Il modello ARIMA(2,0,1) sembra ben adattato ai dati, con coefficienti significativi e valori di **AIC/BIC** piuttosto bassi. Tuttavia, i residui mostrano **eteroschedasticità** e non sono normalmente distribuiti. Questo potrebbe significare che c'è ancora qualche aspetto della serie che non è stato modellato in maniera perfetta, e quindi potrebbe essere utile esplorare modelli GARCH o altre soluzioni che possano modellare la varianza variabile nel tempo.
3. **Predittività**: Il **RMSE** basso suggerisce che il modello è ragionevolmente buono per la previsione

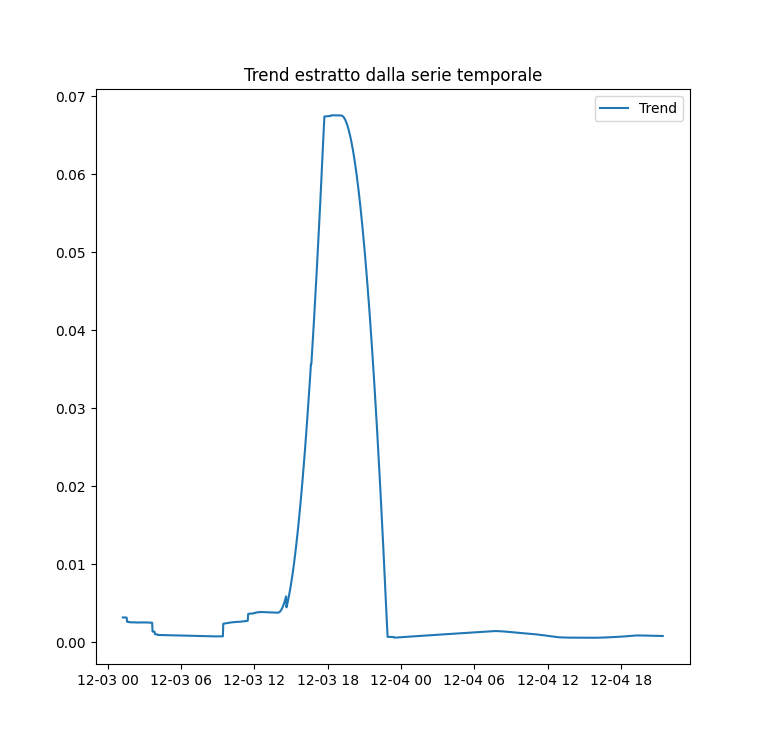
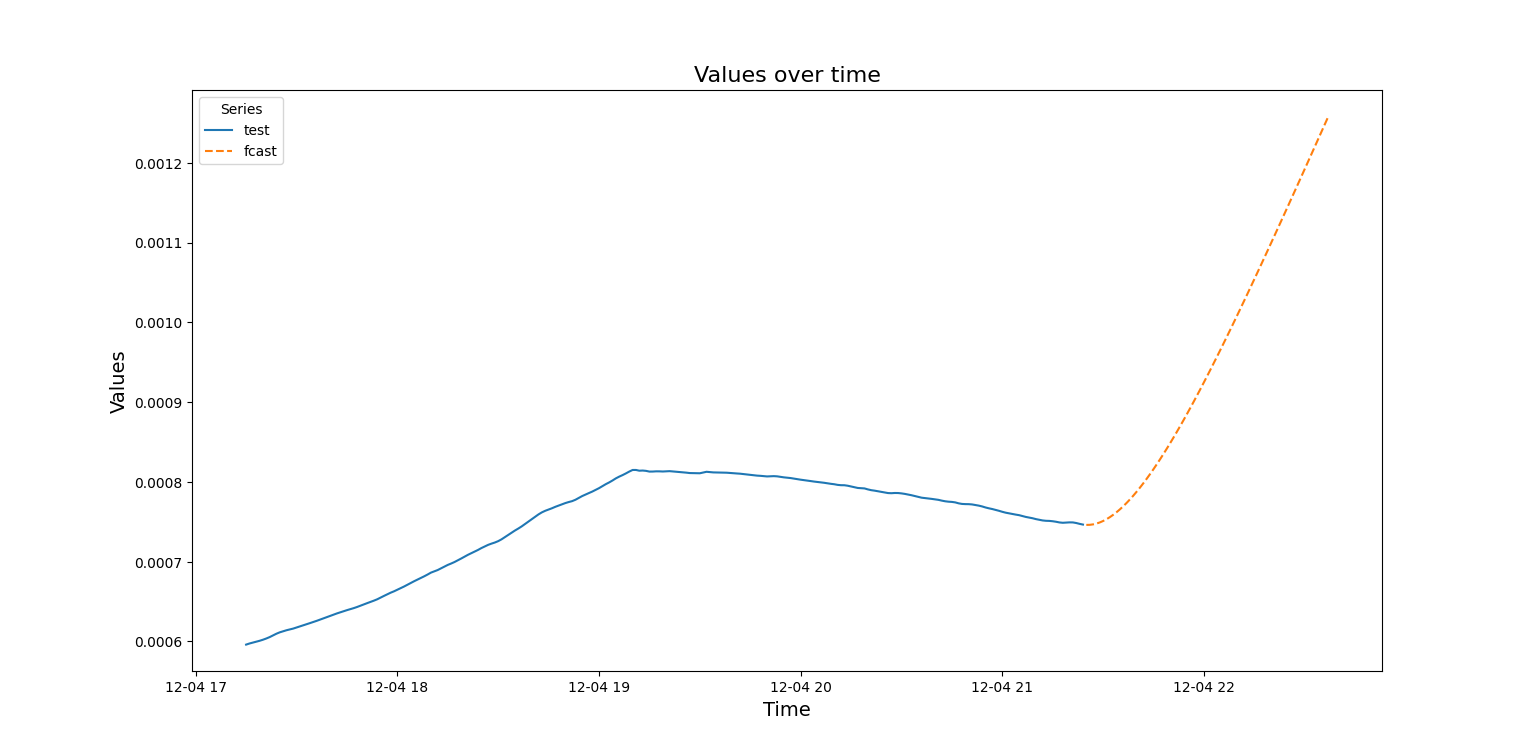
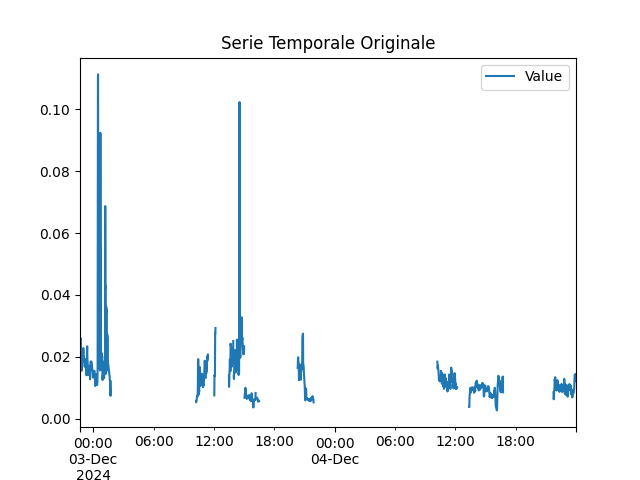


Immagine che contiene testo, diagramma, linea, Diagramma

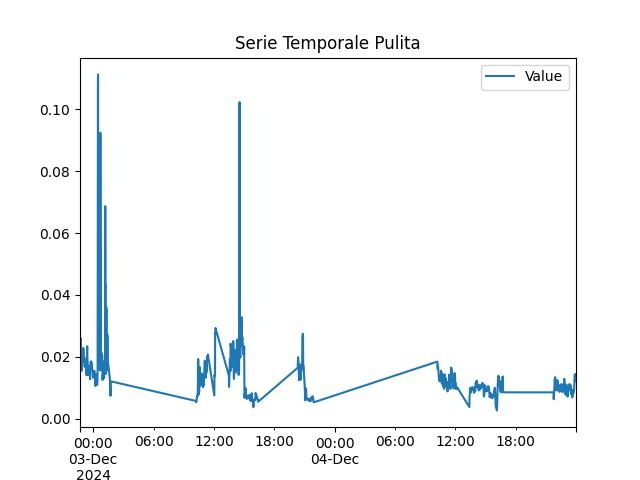
Descrizione generata automaticamente

### 4.1.2 Latenza

Per quel che riguarda la latenza è stato preso in esame un campione di dati raccolti tra il 02/12/2024 e lo 04/12/2024.

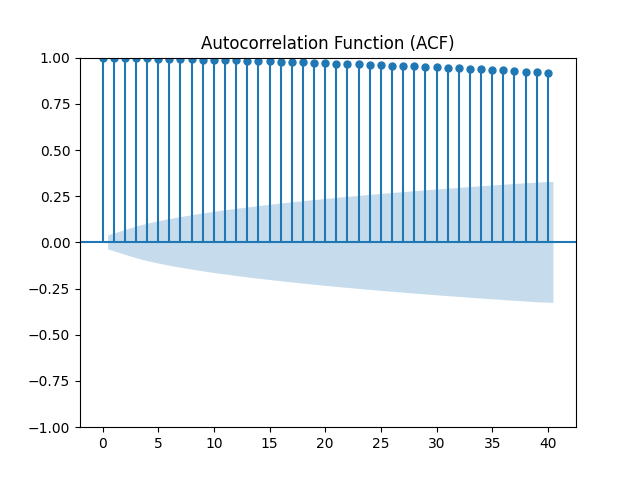


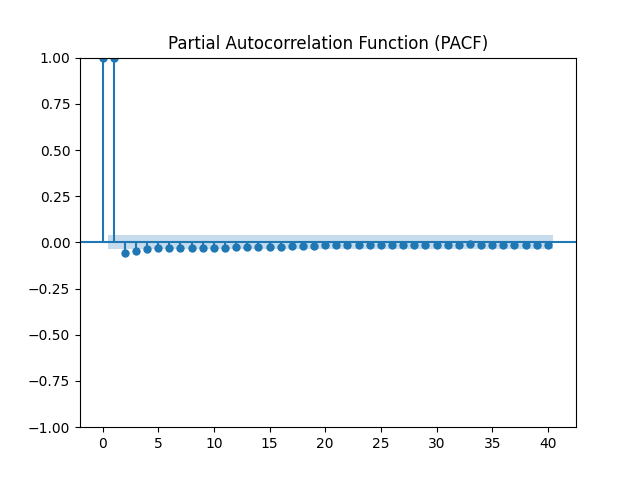
Data l’inconsistenza dei dati è stata applicata interpolazione alla serie originale, il risultato è stato il seguente:



#### 4.1.2.1 Test di Dickey-Fuller Aumentato (ADF)

Il test ADF ha riportato un **test statistic** di -3.018564 e un **p-value** di 0.033197. Poiché il p-value è inferiore a 0.05, possiamo concludere che la serie è stazionaria al livello di significatività del 5%. Questo è importante poiché l'assunzione di stazionarietà è essenziale per i modelli ARIMA.





#### 4.1.2.2 Sommario del Modello ARIMA

* **Modello selezionato**: ARIMA(2, 0, 3)
  + La scelta di **p=2** e **q=3** indica che il modello considera due componenti autoregressive (AR) e tre componenti di media mobile (MA).
  + Non è stata applicata differenziazione (**d=0**), in linea con il fatto che la serie è risultata stazionaria.

**Coefficiente Costante (const)**: 0.0120  
Questo valore rappresenta la media della serie temporale e mostra una stabilità di valore basso.

**Parametri AR (Autoregressivi)**:

* **ar.L1**: 1.9280
* **ar.L2**: -0.9281

I parametri AR catturano la dipendenza lineare tra un punto nella serie temporale e i valori precedenti. Il valore del coefficiente **ar.L1** è positivo e relativamente grande, mentre **ar.L2** è negativo. Questo suggerisce che la relazione tra i valori successivi cambia di segno con il passare del tempo.

**Parametri MA (Media Mobile)**:

* **ma.L1**: 0.8317
* **ma.L2**: -0.1585
* **ma.L3**: -0.3798

I parametri di media mobile rappresentano l'influenza degli errori delle osservazioni precedenti. Questi valori indicano un equilibrio tra effetti positivi e negativi, suggerendo che il modello sta cercando di catturare la volatilità nella serie temporale.

**Varianza dei Residui (sigma2)**: 6.695e-11  
La varianza dei residui è molto piccola, indicando che il modello si adatta bene ai dati e che gli errori sono relativamente limitati.

**Test di Autocorrelazione e Normalità**

* **Ljung-Box (L1) (Q)**: 0.45  
  Questo valore, con un p-value di 0.50, indica che non c'è evidenza significativa di autocorrelazione nei residui, il che è un buon segno per l'adattamento del modello.
* **Jarque-Bera (JB)**: 1061537.25  
  Il test di Jarque-Bera mostra un valore molto elevato con un p-value di 0.00, suggerendo che i residui **non sono normalmente distribuiti**. Questo potrebbe influire sulla precisione delle previsioni e indica che potrebbero esserci alcune strutture nei dati che il modello non sta catturando adeguatamente.
* **Heteroschedasticity (H)**: 0.01  
  Il test di eteroschedasticità ha rilevato una presenza di varianza non costante nei residui (con un p-value di 0.00). L'eteroschedasticità può suggerire che la variabilità dei dati non sia uniforme, e questo può influenzare la robustezza del modello.

**Misure di Errore del Modello**

* **MSE (Mean Squared Error)**: 4.0486e-06
* **RMSE (Root Mean Squared Error)**: 0.0020

L'errore quadratico medio (MSE) e l'errore quadratico medio (RMSE) sono entrambi molto bassi, suggerendo che il modello fornisce previsioni abbastanza precise rispetto ai dati storici. Tuttavia, la normalità dei residui e l'eteroschedasticità devono essere considerate per migliorare ulteriormente la qualità del modello.

**Distribuzione della Serie Temporale**

* **Media**: 0.0114
* **Deviazione standard**: 0.0028
* **Valori estremi**:
  + **Minimo**: 0.0075
  + **Massimo**: 0.0181

I valori indicano che la serie temporale ha una **distribuzione ristretta**, con la maggior parte dei valori compresi tra 0.0075 e 0.0181.

#### 4.1.2.3 Discussione delle Scelte

1. **Ordine del Modello**:
   * La scelta di un modello ARIMA(2,0,3) è stata fatta in base all'analisi della stazionarietà e alla ricerca di un modello che potesse catturare adeguatamente le dinamiche della serie temporale.
   * I due termini autoregressivi (**p=2**) catturano l'influenza dei valori passati, mentre i tre termini di media mobile (**q=3**) cercano di modellare gli shock o gli errori passati.
2. **Stazionarietà**:
   * Poiché il test ADF ha indicato che la serie era già stazionaria, non è stata applicata la differenziazione (**d=0**). Questo aiuta a mantenere il modello più semplice senza perdita di informazioni.
3. **Adattamento del Modello**:
   * I test di Ljung-Box e il valore ridotto di RMSE indicano che il modello si adatta bene alla maggior parte delle caratteristiche della serie temporale.
   * Tuttavia, l'elevato valore del test di Jarque-Bera e la presenza di eteroschedasticità suggeriscono che potrebbe essere necessario considerare ulteriori trasformazioni o modelli alternativi (come un modello GARCH per la varianza non costante).
4. **Warning del Modello**:
   * I warning relativi alla matrice di covarianza "singolare o quasi-singolare" indicano possibili problemi di **multicollinearità** o instabilità dei parametri. Questo suggerisce che, anche se i parametri sono significativi, l'affidabilità degli errori standard può essere compromessa. Potrebbe essere utile riesaminare i dati per verificare se esistono collinearità tra i termini AR e MA.

