



Time Series Approach for Analysis and Prediction of Malware Trends Based on Open Source Intelligence

Tommy Nugraha Manoppo^{1*}, Abdul Gani Fadhlulrahman², Yudi Prayudi³

¹Program Studi Informatika, Fakultas Teknik, Universitas Sulawesi Barat, Indonesia

²Program Studi Teknik Komputer, Fakultas Teknik, Universitas Negeri Gorontalo, Indonesia

³Program Studi Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia, Indonesia

E-Mail: ¹tommynugrahamanoppo@unsulbar.ac.id,

²abdulganiL@ung.ac.id, ³prayudi@uii.ac.id

Received Jan 13th 2026; Revised Mar 06th 2026; Accepted Mar 13th 2026; Available Online Apr 19th 2026

Corresponding Author: Tommy Nugraha Monoppo

Copyright ©2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

The growing threat of malware requires monitoring approaches that provide a continuous, measurable overview of threat trends. This study proposes an open-source intelligence-based malware trend monitoring system using time series forecasting and anomaly signaling. Data are obtained from the MalwareBazaar repository and processed into a daily malware activity time series, with contextual aggregation to identify dominant malware families. The AutoRegressive Integrated Moving Average (ARIMA) model is applied for short-horizon prediction, and statistical anomaly detection is implemented using Z-scores to flag deviations in activity. The system is delivered as an interactive dashboard that visualizes daily malware trends, dominant malware families, forecasting outputs, and anomaly indicators. Experimental results show that ARIMA(2,0,0) provides measurable improvement over a naive persistence baseline, reducing MAE from 102.75 to 92.67 and RMSE from 125.13 to 109.73, while improving sMAPE from 26.74% to 24.48% on the evaluation window. The novelty of this work lies in integrating an OSINT malware repository signal, benchmarked statistical forecasting, quantitative evaluation, and anomaly signaling into a single monitoring dashboard. Practically, the system can support SOC analysts by providing early-warning cues for monitoring prioritization and support digital forensic practitioners by strengthening digital forensic readiness through earlier visibility of emerging malware activity dynamics and dominant artifact categories.

Keywords: ARIMA, Digital Forensic Readiness, Open Source Intelligence, Time Series Forecasting, Malware Trends

1. INTRODUCTION

The rapid evolution of malware remains a persistent challenge for modern cybersecurity because defensive operations increasingly face high-volume submissions, rapid variation, and short decision windows for prioritization and response [1]. In parallel with this escalation, Cyber Threat Intelligence and Open Source Intelligence have become important approaches for obtaining legally accessible threat signals and contextual information to support situational awareness and earlier decision-making [2]. Research on OSINT-driven discovery and interpretation of malicious behavior further indicates that open data can be leveraged to derive meaningful behavioral insights, but its operational value depends on how well the collected signals are structured and interpreted into usable outputs [3]. Threat intelligence ecosystems have also matured through platforms and standards that support collaboration and the structured sharing of indicators; collaborative intelligence-sharing platforms have been studied as practical mechanisms for operationalization across organizations [4]. Nevertheless, the availability of intelligence sources does not automatically translate into operational effectiveness, as integration, data quality, validation, and timeliness remain recurring constraints in generating actionable monitoring outputs [2].

Open malware repositories and community-driven platforms contribute to reproducible security analytics by providing accessible samples, metadata, and automation interfaces that reduce reliance on proprietary telemetry. MalwareBazaar is positioned as a community-driven malware sample exchange and provides a documented interface for automated intelligence consumption and regular access to curated data, enabling structured downstream analytics using an openly accessible repository [5]. Such open repositories are highly relevant to Security Operations Centers, where continuous monitoring, prioritization, and triage require outputs that are measurable and usable within daily workflows [6]. However, many practical uses of

open malware repositories still emphasize descriptive exploration, sample retrieval, or post-incident enrichment, while fewer implementations present end-to-end pipelines that transform open repository signals into predictive monitoring outputs with measurable performance and interpretable decision support for SOC workflows [6]. Despite the availability of open malware repositories and the maturity of time-series methods, reproducible end-to-end implementations that convert open malware signals into evaluated forecasts and operational monitoring outputs remain limited [1], [6].

From a digital forensics perspective, these operational requirements align with Digital Forensic Readiness, which emphasizes preparation before incidents to reduce investigation time and cost while preserving evidence integrity and accountability [7]. Readiness research also highlights that effective implementation requires governance, measurable capability maturation, and dependable evidence-handling procedures rather than ad hoc technical activities [7]. Empirical and case-based studies on digital evidence handling show that organizational procedures and institutional mechanisms strongly influence evidence credibility and the effectiveness of cybercrime case handling, reinforcing the need for structured readiness grounded in operations [8]. A recurring limitation, however, is that readiness implementations often remain focused on post-incident investigation preparedness, while intelligence streams that could support earlier anticipation and evidence preparedness are not consistently integrated into predictive monitoring workflows [7], [8].

Time-series analytics provides a natural bridge toward proactive monitoring because it supports forecasting and deviation detection over continuously collected signals. A recent cybersecurity time-series review reports extensive methodological coverage but also emphasizes gaps in reproducibility, dataset constraints, and the limited availability of consistent end-to-end implementations that connect data acquisition, modeling, evaluation, and interpretation [1]. In security monitoring contexts, ARIMA-based forecasting has also been discussed as a practical approach to monitor security-related metrics over time when evaluation procedures are clearly defined and reported [9]. Complementary survey work on time-series anomaly detection further underscores that practical monitoring requires not only predictive models but also reliable deviation signaling and clear evaluation criteria, especially under volatility and spikes that can degrade naive monitoring assumptions [10]. This work examines whether daily malware sample volume from an openly accessible repository can be forecast with ARIMA and translated into monitoring and early-warning signals with measurable error performance [5], [9]. Practically, the forecasting and anomaly outputs can be used to maintain routine watchlists, trigger triage when deviations occur, and guide short-term monitoring decisions using open-repository trends as a lightweight operational signal [6], [9], [10].

Motivated by this gap, this study develops an OSINT-based malware trend monitoring system that integrates open repository data access, time-series dataset construction from daily sample volume, ARIMA-based forecasting, and statistically grounded anomaly signaling within an interactive dashboard [5], [9], [10]. The contribution lies in an integrated and reproducible pipeline that links open-repository data access, daily-volume time-series construction, ARIMA forecasting, and anomaly signaling into a dashboard-oriented monitoring output [5], [6], [9]. The proposed design emphasizes reproducibility by relying on an openly accessible repository while also reporting quantitative forecasting performance using standard error metrics and presenting outputs in a monitoring-friendly form to support prioritization and early-warning interpretation in SOC contexts [6], [9]. In addition, the monitoring capability is positioned as a practical extension of forensic readiness by treating emerging malware activity dynamics as actionable early signals that can strengthen preparedness for evidence handling and response decision-making beyond post-incident investigation readiness [7], [8]. Daily sample counts are derived from the repository feed, evaluated using a temporal train-test split with error metrics such as MAE and RMSE, and deviations from forecast baselines are flagged as anomalies for dashboard presentation [5], [9], [10]. The literature set was compiled through database searches using predefined keywords and inclusion criteria that prioritized peer-reviewed works on OSINT and CTI, SOC operations, forensic readiness, and time-series forecasting and anomaly detection, with the selection decisions documented for traceability [11]. The expected implication is the availability of an interpretable early-warning view that supports SOC triage prioritization and strengthens evidence-preparedness actions within forensic readiness workflows [6]–[8].

2. MATERIALS AND METHOD

Figure 1 presents the research flow implemented in this study. The workflow consists of OSINT data retrieval from MalwareBazaar, baseline and ARIMA forecasting, quantitative forecast evaluation, anomaly detection using Z-scores, weekly analysis of dominant malware families, and dashboard visualization.

2.1 OSINT Data Retrieval

This study uses MalwareBazaar as an open-source intelligence platform that provides an openly accessible malware repository for legal threat intelligence analysis [5]. Data retrieval is performed programmatically through the MalwareBazaar interface and API to support automated daily collection. The daily malware activity signal is obtained from MalwareBazaar's daily repository archives by counting the

number of entries contained in each daily archive, producing a daily activity sequence for the observation window. In this work, the activity series is constructed over a 14-day observation window with daily granularity to support initial short-horizon monitoring and evaluation. Repository-provided metadata fields, such as malware signature and file name, are also retrieved for descriptive aggregation to support the dominant malware family panel in the dashboard [5].

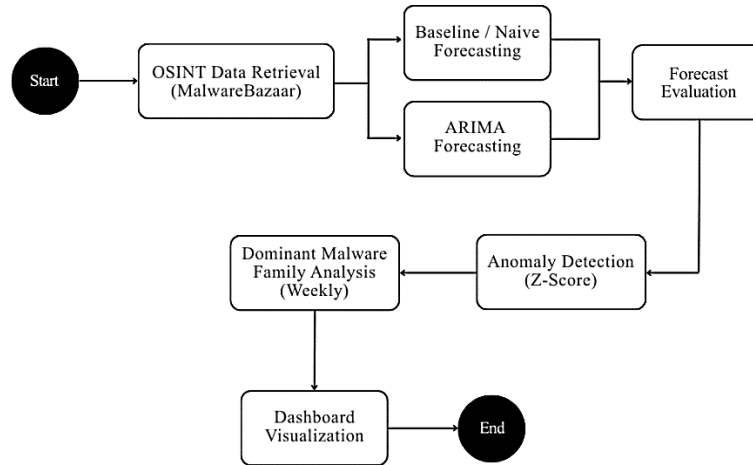


Figure 1. Research Stages of OSINT-Based Malware Trend Analysis System

2.2 Baseline Forecasting

To provide a minimum benchmark for forecasting, this study applies a naive persistence baseline. The baseline predicts the next observation using the previous day's value; see Equation 1.

$$\hat{y}_t = y_{t-1} \quad (1)$$

This baseline provides a transparent reference for short-horizon monitoring and is commonly used in forecasting evaluation as a minimal benchmark [12], [13]. Using a baseline supports interpretation of whether ARIMA provides measurable improvement beyond persistence behavior under volatility [16].

2.3 ARIMA Forecasting

Forecasting is performed using the AutoRegressive Integrated Moving Average model denoted as $ARIMA(p, d, q)$, where p is the autoregressive order, d is the differencing degree, and q is the moving average order [13], [14], [20]. ARIMA is selected because it is a widely adopted statistical approach for short-horizon forecasting and is frequently used as a benchmark model in comparative forecasting studies [17].

Stationarity is assessed using the Augmented Dickey–Fuller test. If the series indicates a unit root, differencing is applied and the test is repeated to determine the differencing degree d [13], [14]. Parameter selection given the short observation window used in this study, ARIMA configurations are evaluated using small candidate orders for stability. Candidate configurations are restricted to $p \in \{0, 1, 2\}$ and $q \in \{0, 1\}$, while d follows the differencing decision from the stationarity procedure. The final configuration is selected based on the lowest forecasting error under the evaluation protocol described in the next subsection, prioritizing MAE and RMSE for selection [12], [13]. Model fitting and forecasting are implemented using a standard ARIMA software interface [15].

Table 1. ADF Stationarity Test Results for Daily Malware Activity Series

Series	ADF Statistic	p -value	Stationary at 5%	Decision
Level y_t	-2.570888	0.099165	No	Apply first differencing
First difference Δy_t	-3.001244	0.034778	Yes	Set $d = 1$

To ensure that the time series satisfies the stationarity requirement for ARIMA modeling, an Augmented Dickey–Fuller (ADF) unit-root test was applied to the daily malware activity series, as summarized in Table 1. The ADF result for the level series indicates that stationarity cannot be confirmed at the 5% significance level (p -value = 0.099165). After first-order differencing, the ADF test indicates that the series is stationary (p -value = 0.034778). Therefore, the differencing degree was set to $d = 1$, and the ARIMA model was constructed using the first-differenced series as the stationary representation for forecasting.

2.4 Forecast Evaluation

Forecast performance is evaluated by comparing predictions from the naive baseline and ARIMA against the actual daily activity values. Let y_t denote the observed daily malware activity and $\hat{y}_t$ denote the forecast at day t . The forecast error is defined as $e_t = y_t - \hat{y}_t$. Quantitative evaluation is reported using MAE, RMSE, MAPE, and sMAPE, which are widely used measures of forecast accuracy [12], [13].

Mean Absolute Error (MAE) measures the average absolute deviation, can view Equation 2.

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (2)$$

Root Mean Squared Error (RMSE) penalizes larger errors more heavily; see Equation 3.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (3)$$

Mean Absolute Percentage Error (MAPE) reports relative error as a percentage of observed values. Equation 4 can be viewed.

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (4)$$

Symmetric Mean Absolute Percentage Error (SMAPE) reduces sensitivity to small denominators; see Equation 5.

$$sMAPE = \frac{100}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{\left(\frac{|y_t| + |\hat{y}_t|}{2} \right)} \quad (5)$$

These metrics are computed for both the naive and the ARIMA models within the same evaluation window, enabling direct performance comparison. The selection and reporting of multiple error measures follow forecast evaluation guidance that emphasizes metric choice and interpretability in volatile signal environments [16]. Since short windows can yield variability in measured performance, results are interpreted as an initial quantitative assessment, consistent with forecasting benchmark observations that emphasize sensitivity to horizon and dataset characteristics [17], [19].

2.5 Anomaly Detection

An anomaly detection mechanism is implemented to identify days where malware activity deviates substantially from normal variation. A Z-score-based approach is used to standardize the daily activity series; see Equation 6.

$$z_t = \frac{y_t - \mu}{\sigma} \quad (6)$$

where μ and σ represent the mean and standard deviation estimated from the available historical window. A day is flagged as anomalous when $|z_t|$ exceeds a predefined threshold. This approach provides an interpretable statistical rule for signaling potential spikes or drops in activity and aligns with general anomaly detection principles for identifying observations that deviate significantly from expected behavior in monitoring settings [18].

2.6 Dominant Malware Family Analysis

To support interpretability, the system aggregates dominant malware categories weekly using repository-provided signature information. Weekly aggregation is used to observe changes in threat composition across consecutive periods, and the resulting summaries are presented in the dashboard as contextual information supporting interpretation of trend, forecasting, and anomaly outputs [5].

2.7 Dashboard Visualization

All outputs are integrated into an interactive dashboard, including daily activity visualization, ARIMA forecasting results, Z-score-based anomaly flags, and weekly summaries of the dominant malware family. The dashboard is designed to support monitoring, prioritization, and early-warning interpretation of malware activity dynamics.

3. RESULTS AND DISCUSSION

This section presents the results of the implementation of an open source intelligence-based malware trend monitoring system as well as a discussion of the findings obtained from the analysis dashboard. The study focuses on visualizing malware activity trends, identifying dominant malware families, and evaluating the system's ability to predict and detect anomalies.

3.1. Implementation of the Malware Monitoring Dashboard

The developed system successfully integrates malware data retrieval, pre-processing, time-series analysis, and result visualization in an interactive dashboard. This dashboard is designed to provide a temporal overview of malware activity conditions, while providing features of dominant malware family analysis, trend prediction, and anomaly detection. The main dashboard view shows the number of daily malware sample submissions as a graph. This visualization represents the dynamics of malware activity in the observation period and serves as the basis for trend analysis and prediction.

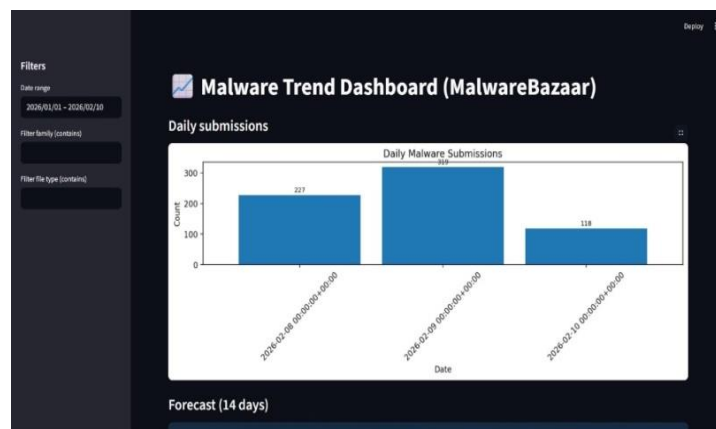


Figure 2. Daily malware delivery trends dashboard view

Based on Figure 2, the number of malware sample shipments in the initial period of February 2026 fluctuated significantly. At one observation point, the number of samples reached more than 300 per day, while on another day it decreased to close to 100. This pattern suggests that malware activity is dynamic and non-constant, requiring a monitoring system capable of conducting continuous analysis. These fluctuations can be caused by a variety of factors, such as malware distribution campaigns, botnet activity, or the spread of new malware variants. Therefore, visualizing daily trends is an important component for detecting changes in the pattern of cyber threat activity.

3.2. Analysis of the Dominant Malware Family

In addition to displaying sample count trends, the system also aggregates and visualizes the most dominant malware families in a given period. This analysis is conducted weekly to monitor changes in the distribution of the malware family.

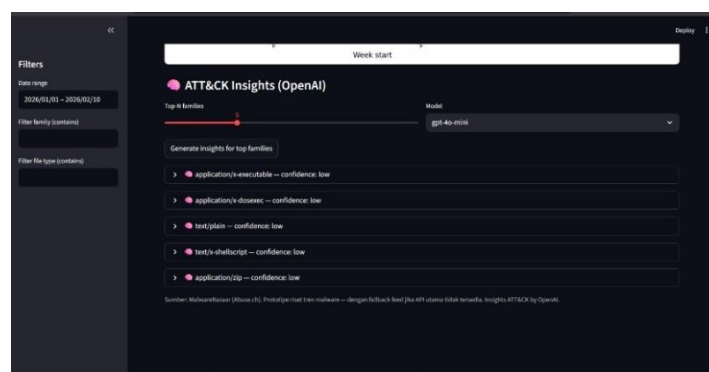


Figure 3. Visualization of dominant malware families on a weekly basis

Based on Figure 3, the executable and executable file categories accounted for the majority of malware samples during the observation period. In one week, the number of DOS executable samples reached more than 180 entries, while the other executable categories also showed significant numbers. The

dominance of the executable file category suggests that program-based malware distribution methods are still the most common technique used by threat actors. In addition, other categories, such as text files, shell scripts, and compressed archives, appear as distribution media, though in fewer numbers. The distribution of this malware family indicates that threat actors are still using various types of files as a means of spread, both as direct payloads and as malware wrappers. This information can be used by security analysts to understand distribution patterns and determine detection priorities

3.3. Malware Activity Prediction, Evaluation and Anomaly Detection

The developed system provides malware activity prediction using an ARIMA-based time series model and compares its predictive performance against a naive persistence baseline. In addition, the system implements an anomaly detection mechanism using a Z-score-based statistical approach to identify activity deviations from normal patterns.

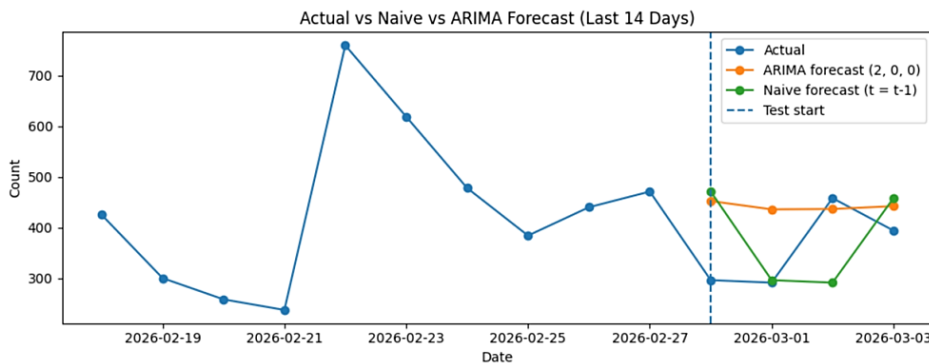


Figure 4. Prediction Results with Naive and ARIMA Forecast

Figure 4 presents a comparison of short-horizon malware activity forecasting using an ARIMA model and a naive persistence baseline against the actual daily activity values. The blue line represents the observed daily malware activity, the orange line represents the ARIMA forecast, and the green line represents the naive forecast where the prediction for day t is set to the observation at day $t - 1$. The vertical dashed line indicates the start of the evaluation window used to assess forecasting performance. The naive baseline tends to follow the previous-day value and therefore reacts strongly to day-to-day fluctuations, which can lead to larger deviations when the activity changes abruptly. In contrast, the ARIMA(2,0,0) forecast provides a more stable short-horizon estimate that better approximates the observed activity level in the evaluation window.

Table 2. Quantitative Forecasting Performance for Malware Activity Prediction

Model	MAE	RMSE	MAPE (%)	sMAPE (%)
ARIMA(2,0,0)	92.67	109.73	29.87	24.48
Naive baseline ($t = t - 1$)	102.75	125.13	28.39	26.74

Quantitative evaluation confirms that the ARIMA(2,0,0) model provides improved forecasting performance compared with the naive persistence baseline, as summarized in Table 2. ARIMA reduces the average absolute error from 102.75 to 92.67 and lowers RMSE from 125.13 to 109.73, indicating that its predictions are closer to the observed daily activity and less affected by large deviations. In relative terms, the sMAPE decreases from 26.74% to 24.48%, showing that ARIMA achieves more consistent percentage-based accuracy across the evaluation window. While MAPE for ARIMA is slightly higher than the baseline, this difference does not outweigh the improvements shown by MAE, RMSE, and sMAPE, which collectively indicate a more reliable short-horizon estimate for daily malware activity. Overall, these results demonstrate that ARIMA delivers a measurable gain over the naive benchmark and therefore can be adopted as an early-indicator component in the malware monitoring dashboard to support routine monitoring and short-term situational awareness.

Based on Figure 5 in the anomaly detection section, the system did not detect any significant deviation during the observation period. This suggests that malware activity during that period remained within the normal range of variation based on the available data. Nonetheless, anomaly detection mechanisms still play an important role in long-term implementation. When there is an unusual spike in malware activity, the system can provide early warning to security analysts, enabling faster responses. Figure 5 also shows a panel of dominant malware families grouped by weekly activity. The graph shows the distribution of the five categories of malware families with the highest frequency over the two observation periods. It can be seen that the *application/x-dosexec* and *application/x-executable* categories dominated activity in both weeks,

with a higher number of samples in the second period. In addition, other categories such as *text/plain*, *application/zip*, and *text/x-shellscript* also appear in smaller numbers, indicating variations in the file types detected in malware activity. This pattern shows that malware activity is not concentrated in a single file type but is spread across several categories, so a weekly malware family analysis can provide an overview of the composition of emerging threats.



Figure 5. Malware Grouping Panel Based on malware activity and anomaly detection

3.4. Discussion

The dashboard results indicate that an open-source intelligence approach can be operationalized to support near-real-time monitoring of malware activity trends by integrating daily activity visualization, forecasting, anomaly signaling, and threat context summaries within a single interface. The observed fluctuations in daily malware activity reflect the fact that the threat ecosystem is not constant and can change rapidly, consistent with the temporal nature of attack campaign cycles, where adversary activity may escalate during the delivery and execution phases and then decay as infrastructure is rotated or disrupted [21].

From a threat intelligence perspective, the developed system can be interpreted within the threat intelligence lifecycle, where collected signals are processed and transformed into analytic outputs that support dissemination and operational consumption by analysts [22], [23]. This lifecycle-oriented interpretation is important because OSINT-derived signals provide greater value when translated into decision-support outputs that can inform monitoring priorities and triage. In this study, the dashboard provides such outputs in the form of daily trend visibility, short-horizon forecasting, and anomaly flags, enabling a monitoring workflow in which deviations can be reviewed and related to dominant malware categories in the same period [22], [23].

The dominant malware family analysis further strengthens operational interpretation of trend dynamics. The observed dominance of executable-based categories indicates that executable delivery remains a relevant vector in real-world incidents, aligning with industry-scale breach analyses that continue to report malware execution, including the execution of attachments, as a recurring contributor in malware-related interactions [24]. This suggests that weekly context summaries can complement daily volume monitoring by indicating whether the dominant delivery artifacts remain stable or shift over time, which is useful for prioritizing detection focus and contextual investigation in operational settings [24].

Regarding forecasting, the quantitative comparison between ARIMA and the naive baseline provides evidence that statistical time-series forecasting can deliver measurable improvements for short-horizon prediction. In the evaluation window, ARIMA achieves lower magnitude-based errors than the naive baseline and improves relative error stability as reflected by sMAPE, indicating that the forecasting module provides added value beyond a persistence assumption. However, the forecasting behavior remains sensitive to volatility and sudden turning points, which is expected given that ARIMA is fundamentally a linear model intended to capture autocorrelation and trend components and may not fully represent non-linear dynamics driven by abrupt campaign transitions or bursty propagation [18]. These limitations justify future development toward hybrid forecasting methods that combine statistical structure with non-linear learning components, which have been widely explored to improve forecasting robustness when time series exhibit non-linearities or regime changes [25].

Overall, the study demonstrates that OSINT-derived malware repository signals can be transformed into a monitoring-oriented dashboard that supports both descriptive and predictive perspectives of malware activity. By interpreting daily and weekly patterns through the lens of campaign dynamics and the threat intelligence lifecycle, the findings provide a stronger theoretical grounding for operational use, while the observed limitations under volatile and non-linear behavior motivate future work toward longer historical collection and hybrid modelling for improved early-warning reliability [21]–[25].

4. CONCLUSION

This study develops an open source intelligence-based malware trend monitoring system that integrates automated data retrieval, daily activity aggregation, ARIMA-based forecasting, Z-score anomaly signaling, and interactive dashboard visualization to support continuous observation of malware activity dynamics, dominant malware family context, and short-horizon prediction under limited historical data. Beyond descriptive monitoring, the system has clear practical implications for digital forensic readiness by enabling earlier situational awareness of emerging malware activity patterns, supporting faster initial triage, and improving evidence-preparedness decisions. In a digital forensics context, the trend and dominant-family panels can assist investigators in prioritizing artifact preservation and preliminary scoping by indicating when activity increases and which file categories or signatures are most prevalent, thereby helping align early collection and contextual interpretation with likely active attack behaviors. From an operational perspective, the dashboard is relevant to SOC environments because it provides a lightweight monitoring layer that complements routine workflows, such as prioritization, watchlist updates, and early-warning review when abnormal deviations are detected.

Future work should expand the historical collection window and evaluate performance over longer, more diverse observation periods to strengthen forecasting stability and anomaly sensitivity. In addition to ARIMA, deep learning time-series models and hybrid approaches should be explored to better handle non-linear and bursty patterns in malware activity. Further improvements include integrating external threat intelligence sources to enrich context and correlation, and mapping dominant families and observed artifacts to the MITRE ATT&CK framework to support more actionable interpretation for both SOC analysts and forensic practitioners. Finally, real-world deployment evaluations in operational SOC settings and digital forensic laboratories are recommended to assess practical effectiveness, including usability, alert usefulness, integration feasibility, and measurable impacts on triage speed, evidence-handling readiness, and investigative outcomes.

ACKNOWLEDGMENTS

The author would like to thank all parties who have helped in the compilation of this work. Thank you to the supervisor for the guidance and direction given, to the institution where the research was conducted for permission and data support, and to all respondents who have participated. The author also thanks friends and family for the prayers, support, and motivation given during the process of completing this research

REFERENCES

- [1] M. Landauer, F. Skopik, B. Stojanović, A. Flatscher, and T. Ullrich, "A review of time-series analysis for cyber security analytics: from intrusion detection to attack prediction," *International Journal of Information Security*, vol. 24, no. 1, Art. no. 3, 2024, doi: 10.1007/s10207-024-00921-0.
- [2] E. Condon, A. He, and M. Cukier, "Analysis of Computer Security Incident Data Using Time Series Models," in *Proc. IEEE International Symposium on Software Reliability Engineering (ISSRE)*, 2008, pp. 77–86, doi: 10.1109/ISSRE.2008.39.
- [3] W. Tounsi and H. Rais, "A survey on technical threat intelligence in the age of sophisticated cyber attacks," *Computers & Security*, vol. 72, pp. 212–233, 2018, doi: 10.1016/j.cose.2017.09.001.
- [4] Y.-T. Huang, C.-Y. Lin, Y.-R. Guo, K.-C. Lo, Y. S. Sun, and M.-C. Chen, "Open Source Intelligence for Malicious Behavior Discovery and Interpretation," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 2, pp. 776–789, 2022, doi: 10.1109/TDSC.2021.3119008.
- [5] abuse.ch, "MalwareBazaar: Malware sample exchange platform," 2026. [Online]. Available: bazaar.abuse.ch. Accessed: Mar. 5, 2026.
- [6] C. Wagner, A. Dulaunoy, G. Wagener, and A. Iklody, "MISP: The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform," in *Proc. ACM Workshop on Information Sharing and Collaborative Security (WISCS)*, 2016, pp. 49–56, doi: 10.1145/2994539.2994542.
- [7] M. Vielberth, F. Böhm, I. Fichtinger, and G. Pernul, "Security Operations Center: A Systematic Study and Open Challenges," *IEEE Access*, vol. 8, pp. 7756–7782, 2020, doi: 10.1109/ACCESS.2020.3045514.
- [8] A. Taiwo, F. Bankole, and I. Claims, "An extended digital forensic readiness and maturity model," *Forensic Science International: Digital Investigation*, vol. 40, Art. no. 301348, 2022, doi: 10.1016/j.fsidi.2022.301348.
- [9] Y. Prayudi, A. Ashari, and T. K. Priyambodo, "The Framework to Support the Digital Evidence Handling: A Case Study of Procedures for the Management of Evidence in Indonesia," *Journal of Cases on Information Technology*, vol. 22, no. 3, 2020, doi: 10.4018/JCIT.2020070104.
- [10] J. Kohlrausch and E. A. Brin, "ARIMA Supplemented Security Metrics for Quality Assurance and Situational Awareness," *Digital Threats: Research and Practice*, vol. 1, no. 1, 2020, doi: 10.1145/3376926.

-
- [11] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, 2009, doi: 10.1145/1541880.1541882.
 - [12] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006, doi: 10.1016/j.ijforecast.2006.03.001.
 - [13] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021.
 - [14] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.
 - [15] Statsmodels Developers, "ARIMA model (statsmodels.tsa.arima.model.ARIMA) documentation," accessed Mar. 5, 2026.
 - [16] H. Hewamalage, K. Ackermann, and C. Bergmeir, "Forecast evaluation for data scientists: common pitfalls and best practices," *Data Mining and Knowledge Discovery*, 2023, doi: 10.1007/s10618-022-00894-5.
 - [17] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLOS ONE*, vol. 13, no. 3, e0194889, 2018, doi: 10.1371/journal.pone.0194889.
 - [18] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 Competition: 100,000 time series and 61 forecasting methods," *International Journal of Forecasting*, vol. 36, no. 1, pp. 54–74, 2020, doi: 10.1016/j.ijforecast.2019.04.014.
 - [19] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. C. Aggarwal, and M. Salehi, "Deep Learning for Time Series Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 57, no. 1, Art. no. 15, 2024, doi: 10.1145/3691338.
 - [20] C. Johnson et al., "Guide to Cyber Threat Information Sharing," NIST Special Publication 800-150, 2016.
 - [21] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains," Lockheed Martin white paper, 2011.
 - [22] FIRST CTI SIG, "Methods and Methodology: Cyber Threat Intelligence cycle and workflow," accessed Mar. 5, 2026.
 - [23] C. Johnson et al., "Guide to Cyber Threat Information Sharing," NIST SP 800-150, accessed Mar. 5, 2026.
 - [24] Verizon, 2025 Data Breach Investigations Report (DBIR), 2025.
 - [25] S. Baratsas et al., "A hybrid statistical and machine learning based forecasting framework for the energy sector," *Computers & Chemical Engineering*, 2024, doi: 10.1016/j.compchemeng.2024.108740.
-