DESIGNING A SYSTEM TO PREDICT FACULTY PUBLICATION PERFORMANCE USING A TIME-SERIES MACHINE LEARNING MODEL (CASE STUDY: INFORMATION TECHNOLOGY STUDY PROGRAM, PGRI UNIVERSITY OF MADIUN)

Wilhelmus Hia[1)], Puguh Jayadi[2)], Yessi Yunitasari[*,3)]

[1,2,3]Department of Informatic Engineering, Universitas PGRI Madiun, Indonesia
85 Setia Budi Street, Madiun City, East Java, Indonesia
Email: [1]wilhelmus_2205101116p@mhs.unipma.ac.id , [2]puguh.jayadi@unipma.ac.id , [3*]yessi@unipma.ac.id

ABSTRACT

This research presents the development of an integrated web-based application designed to forecast and assess lecturers' publication performance through time-series machine learning techniques. The study is conducted within the Informatics Engineering Program of Universitas PGRI Madiun by utilizing historical publication records as temporally ordered data. These records are transformed into structured time-series representations to reveal productivity trends, periodic patterns, and temporal variations in academic output. The system architecture consolidates data acquisition, preprocessing, forecasting, and visualization into a single platform implemented using the Django framework. Multiple time-series forecasting models are applied and evaluated to examine their predictive capability, with performance measured using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The results indicate that the proposed system is capable of generating consistent and interpretable forecasts that reflect actual publication dynamics. By providing analytical insights and early trend detection, the system supports objective evaluation and assists academic management in strategic planning and continuous improvement of research productivity.

Keywords: Time-Series Machine Learning, Publication Performance, Web-Based System, Academic Evaluation, Decision Support System

1. INTRODUCTION

The evaluation of lecturers' academic performance plays a strategic role in maintaining the quality and sustainability of higher education institutions [1]. Research productivity, particularly in the form of scientific publications, is widely regarded as a tangible representation of academic contribution and institutional reputation [3]. Unlike subjective assessment criteria, publication output can be systematically verified through bibliometric records, including indexed journals, citation profiles, and publication frequency across recognized platforms [7]. Consequently, publication performance has become a central reference in institutional evaluation frameworks and policy formulation [2].

Despite its importance, publication performance assessment in many universities is still conducted through manual recapitulation and retrospective reporting [6]. These practices often depend on fragmented data sources and static summaries, which limits their ability to reflect temporal changes in research productivity [4]. As publication activity is influenced by various factors—such as academic calendars, research funding cycles, and institutional policies—static evaluation approaches are insufficient to capture fluctuations and emerging trends over time [9]. This limitation reduces the usefulness of evaluation results for forward-looking planning and early intervention [5].

Modeling publication records as time-dependent data provides an opportunity to address these challenges [8]. Time-series analysis enables historical publication patterns to be examined as sequential observations, allowing both trend identification and future projection [11]. Recent advances in machine learning have further strengthened time-series forecasting by improving the ability to capture temporal dependencies and recurring patterns that are often overlooked by conventional descriptive statistics [14]. When applied appropriately, these techniques can generate forecasts that reflect realistic academic productivity trajectories [10].

From a systems perspective, predictive models alone are not sufficient to support institutional decision-making [12]. Without integration into an operational platform, analytical results remain difficult to access, interpret, and utilize by academic stakeholders [15]. Web-based systems offer a practical solution by providing centralized data management, automated processing, and real-time access to evaluation outcomes [18]. The integration of predictive analytics into such platforms transforms performance assessment from a passive reporting activity into an active monitoring process that supports evidence-based management [13].

Although numerous studies have explored bibliometric analysis and forecasting techniques, many of them emphasize methodological accuracy without addressing system-level implementation [16]. As a result, the practical adoption of predictive analytics in academic performance management remains limited [19]. There is a need for solutions that not only deliver reliable forecasts but also embed them within usable systems that translate analytical outputs into actionable insights [17].

In response to this gap, this study develops a web-based system that integrates time-series machine learning models to predict and evaluate lecturers' publication performance within the Informatics Engineering Program of Universitas PGRI Madiun [20]. By combining forecasting models, standardized evaluation metrics, and interactive visualization, the proposed system is designed to support transparent assessment, early detection of performance trends, and informed strategic planning in higher education research management [21].

2. METHOD

This study adopts a systematic methodological approach to develop a time-series–based forecasting system for evaluating lecturers' publication performance. The research process begins with an analysis of institutional publication records to identify relevant evaluation requirements, temporal structures, and seasonal characteristics of academic productivity. Historical publication data are then organized into sequential time-series formats and modeled using the Autoregressive Integrated Moving Average (ARIMA) method to capture non-seasonal trends, followed by the Seasonal ARIMA (SARIMA) model to account for recurring patterns associated with academic cycles. Model training and validation are conducted using historical data, with predictive performance assessed through Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2) to ensure reliability and consistency of the forecasting results, the following is the implementation flow carried out by the researcher:

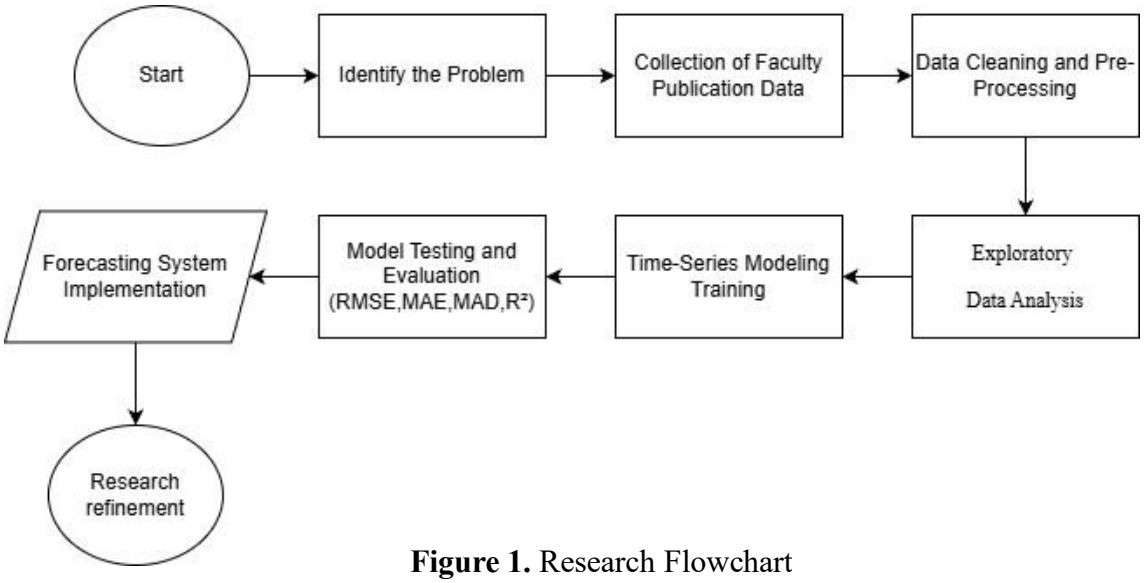

Figure 1. Research Flowchart

1. Dataset Description

The data utilized in this study are derived from lecturers' publication records maintained by the Informatics Engineering Program of Universitas PGRI Madiun and verified through established academic indexing services, including Google Scholar, Scopus, and SINTA. These records are compiled over a specified observation period and reorganized into a chronological structure to support time-series analysis. Publication outputs are aggregated at a uniform temporal interval, either annually or per academic semester, to preserve consistency across analytical stages. Each temporal unit represents the total number of publications and is supplemented with descriptive attributes such as publication period, output classification, and indexing status. Before the forecasting models are applied, the dataset is systematically preprocessed through data screening, duplication checks, and the treatment of incomplete records to improve data quality. This structured and validated dataset serves as the empirical basis for examining publication trends, seasonal behavior, and temporal variability using ARIMA and SARIMA modeling approaches.

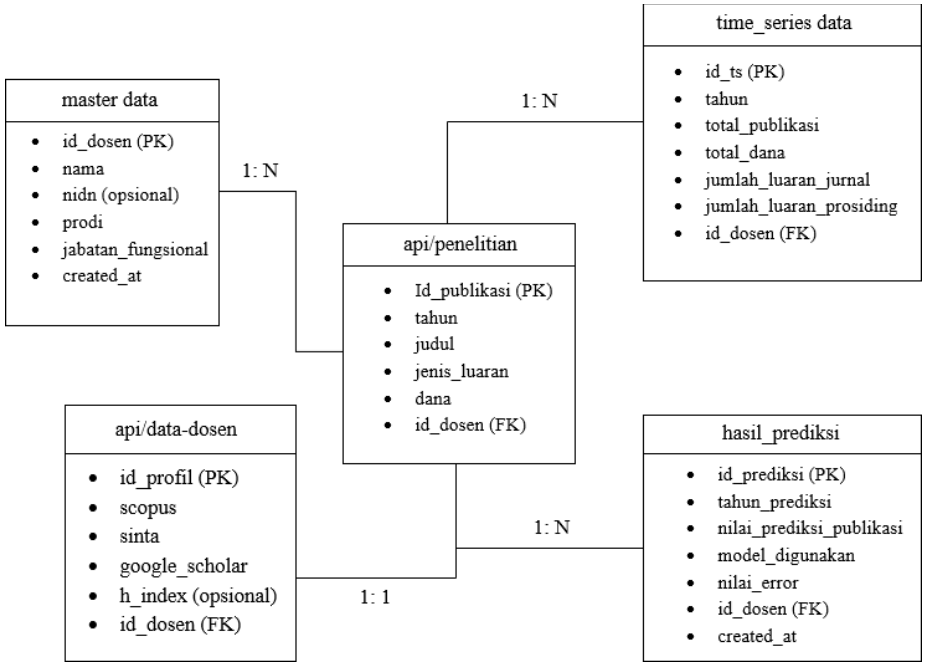

Figure 2. ERD Database

Figure ERD illustrates the logical database design that supports the structured time-series dataset described previously by organizing lecturers' publication data, indexing profiles, and prediction outputs in an integrated manner. The master data entity serves as the core reference for lecturer identities and academic attributes, which are linked in a one-to-many relationship with the api/penelitian entity that stores detailed publication records by year, output type, and research funding. These publication records are further aggregated into the time_series_data entity to form annual or semester-based summaries, providing the temporal structure required for ARIMA and SARIMA modeling. Complementary bibliometric information from external indexing platforms such as Google Scholar, Scopus, and SINTA is managed through the api/data_dosen entity in a one-to-one relationship with lecturer data, ensuring consistency of indexing attributes. Finally, the hasil_prediksi entity records forecasting results, including predicted publication values, applied models, and error metrics, thereby linking analytical outcomes back to individual lecturers. This ERD design ensures data integrity, traceability, and coherence between raw publication data, aggregated time-series inputs, and predictive results used in the proposed system.

2. Variables and Indicators

The variables and indicators in this study are defined based on structured time-series datasets obtained from lecturer publication records, as described in the Dataset Description. The main variables are designed to represent aspects of academic productivity that can be measured over time, ensuring suitability for the purpose of predicting publication performance using ARIMA and SARIMA models. Each variable is selected to capture both the volume and characteristics of scientific output within a consistent time frame, thereby supporting reliable trend and seasonal analysis.

The main dependent variable in this study is the total number of publications per lecturer within a given time period, which serves as the primary indicator of publication performance. This variable is constructed through the temporal aggregation of individual publication records and reflects overall research productivity. Supporting indicators include year of publication, type of output (journal or proceedings), and indexing category, which provide the contextual information needed to interpret variations and patterns in publication output across different periods.

Independent variables and descriptive indicators are obtained from bibliometric and institutional attributes associated with each lecturer. These include indexing profiles from Google Scholar, Scopus, and SINTA, as well as optional indicators such as h-index and research funding allocation. Although these indicators are not directly modelled as predictors.

3. Time-Series Model Configuration

Time-series configuration is a critical stage in ensuring that the predictive model accurately captures the temporal dynamics of lecturers' publication performance. This stage focuses on analyzing the underlying statistical characteristics of the data to determine appropriate model parameters before the forecasting process is conducted.

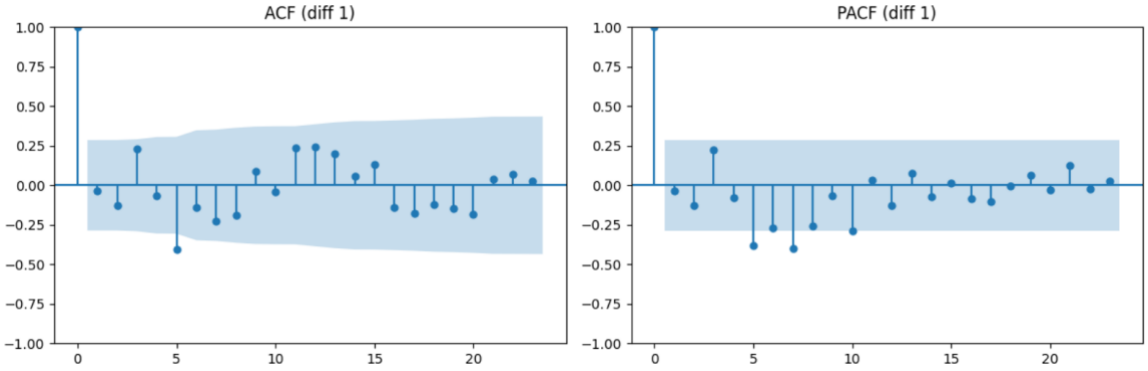

Figure 3. ACF and PACF Model

Figure 3. presents the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the publication time-series after first-order differencing. The differenced series exhibits significantly reduced autocorrelation, indicating that the non-stationary trend has been effectively removed. Most lag values fall within the confidence bounds, suggesting a stable temporal structure with limited residual dependency. Several significant spikes at early lags provide empirical guidance for identifying the autoregressive and moving average components used in configuring the ARIMA and SARIMA models. This analysis supports the selection of appropriate model parameters and ensures that the forecasting process captures meaningful temporal dependencies while maintaining statistical robustness.

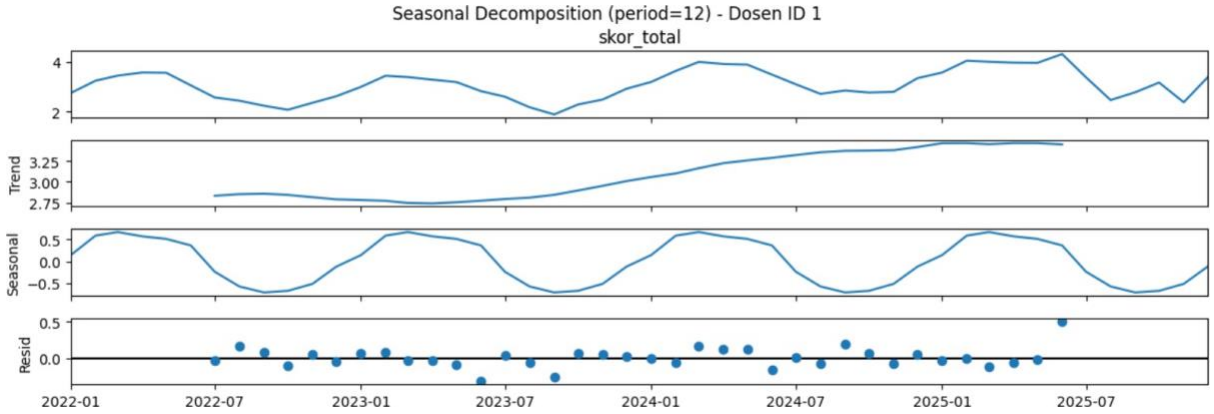

Figure 4. Seasonal Decomposition

Complementing the autocorrelation analysis, Figure Y illustrates the seasonal decomposition of the publication time-series into trend, seasonal, and residual components using a 12-period cycle. The trend component reveals a gradual upward pattern in lecturers' publication performance over time, while the seasonal component exhibits recurring fluctuations that indicate the presence of periodic academic productivity patterns. Meanwhile, the residual component remains relatively small and randomly distributed around zero, suggesting that most systematic temporal variations have been adequately captured by the trend and seasonal structures. This decomposition further validates the suitability of incorporating both non-seasonal and seasonal elements in the ARIMA and SARIMA configurations, reinforcing the robustness of the forecasting framework. Furthermore, the clear separation between trend and seasonal components enables more interpretable modeling of long-term productivity growth and recurring academic cycles. By isolating these components, the forecasting model can reduce noise interference and improve generalization when predicting future publication performance.

4. Model Performance Evaluation

Model performance evaluation was conducted to assess the accuracy and reliability of the proposed time-series forecasting approaches. The predictive results were evaluated using MAE, RMSE, and R^2 to capture both absolute error magnitude and overall explanatory power. Comparative analysis indicates that the seasonal model achieves lower forecasting errors and better goodness-of-fit, highlighting its effectiveness in modeling recurring academic productivity patterns. Furthermore, the alignment between predicted and actual publication trends demonstrates that the proposed system can generate reliable forecasts suitable for supporting data-driven academic performance monitoring.

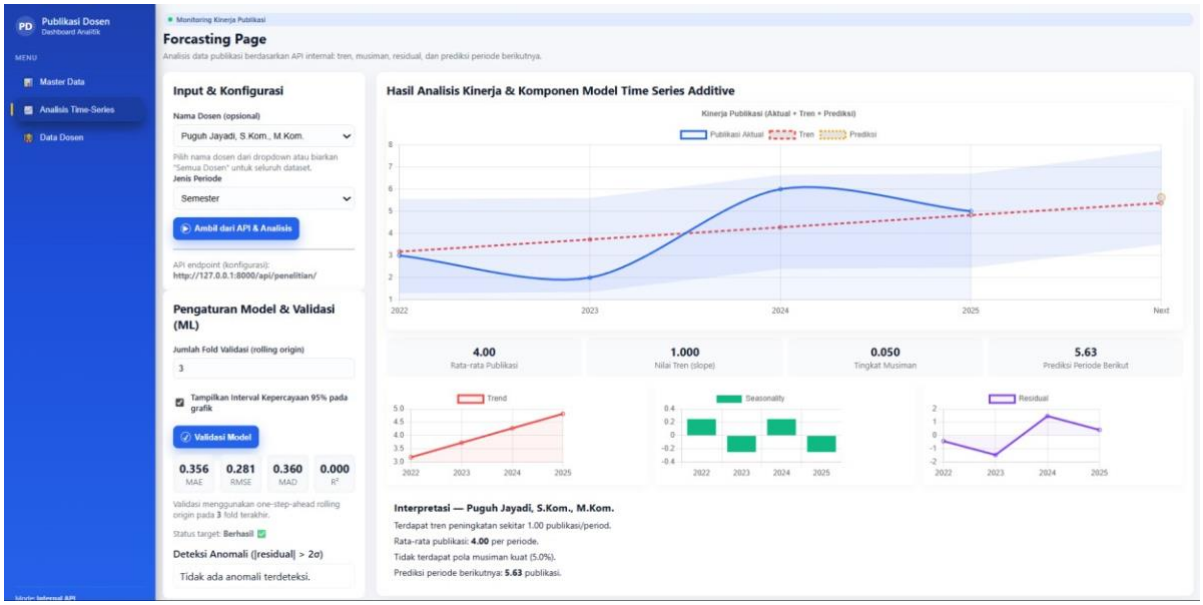

Figure 5. time-series analysis page

Figure 5. illustrates the time-series analysis and forecasting interface implemented in the proposed system, which serves as an integral component of the model performance evaluation process. The dashboard visualizes the alignment between actual publication data, estimated trend, and forecasted values, enabling direct qualitative assessment of the model's predictive behavior. The inclusion of confidence intervals further supports reliability analysis by indicating the uncertainty range of predictions, while the decomposition into trend, seasonality, and residual components provides interpretability regarding how temporal patterns are captured by the model.

In addition to visual evaluation, the system integrates quantitative validation results obtained through rolling-origin cross-validation, as reflected by the MAE, RMSE, MAD, and R^2 values displayed on the analysis page. These metrics allow academic managers and system users to assess prediction accuracy in an intuitive yet statistically grounded manner. By combining numerical performance indicators with interactive visual analytics, the proposed system ensures that forecasting results are not only accurate but also transparent and actionable, reinforcing its role as a practical decision-support tool for monitoring lecturers' publication performance.

5. System Functional Features

The proposed system implements a set of functional features designed to operationalize time-series forecasting within an integrated web-based environment. Publication data are automatically retrieved from an internal API and preprocessed into structured time-series formats, enabling consistent temporal analysis across different lecturers and periods. The system allows configurable model execution, including period selection and rolling-origin validation, ensuring that predictive results are both reproducible and statistically reliable.

Furthermore, the system provides interactive visualization and performance reporting to support interpretability and decision-making. Forecast results are presented alongside trend, seasonal, and residual components, complemented by quantitative accuracy metrics such as MAE, RMSE, MAD, and R^2. Anomaly detection based on residual analysis is also integrated to highlight unusual publication patterns. These features collectively enable academic managers to monitor publication performance, evaluate forecasting reliability, and utilize predictive insights effectively within a single decision-support platform. By embedding analytical, validation, and visualization processes within a unified workflow, the system reduces manual intervention and

minimizes subjectivity in performance assessment. As a result, the proposed platform not only enhances forecasting efficiency but also supports evidence-based academic management through timely and data-driven insights.

3. System Execution Results

The system execution results demonstrate that the proposed web-based forecasting system operates successfully in generating publication performance predictions based on time-series data. Through an integrated workflow consisting of data preprocessing, model execution, validation, and visualization, the system is able to produce consistent forecasting outputs without manual intervention. This confirms that the designed system functions as intended and can reliably execute the complete prediction pipeline in an operational environment. The accuracy of the forecasting results is quantitatively reflected through evaluation metrics including RMSE, MAE, MAD, and the coefficient of determination (R^2). The obtained error values remain close to zero, while the R^2 values approach one, indicating that the predicted publication trends closely follow the actual observed data. These results suggest that the system is capable of minimizing prediction errors while preserving a high level of explanatory power.

Furthermore, the consistency of the error metrics across evaluation periods indicates that the forecasting process is stable and not overly sensitive to short-term fluctuations in the data. Low dispersion in RMSE, MAE, and MAD values reflects controlled residual behavior, while consistently high R^2 values confirm that the model maintains reliable performance over time. This stability is critical for ensuring that the system can be trusted for continuous monitoring of lecturers' publication performance.

Overall, these system execution results validate that the proposed forecasting system delivers accurate, stable, and interpretable predictions suitable for practical use. By combining low prediction errors with strong goodness-of-fit indicators, the system provides a robust foundation for data-driven academic performance evaluation. Consequently, the implemented system can effectively support academic management in identifying publication trends, assessing performance consistency, and making informed strategic decisions.

4. Conclusion

This research concludes that the developed web-based application is effective in supporting the prediction and monitoring of lecturers' publication performance through integrated data management and time-series forecasting. By automating the processing and visualization of publication records, the system eliminates reliance on manual recapitulation and provides a consistent mechanism for tracking research productivity. The forecasting component exhibits dependable predictive behavior, reflected by relatively low error values (MAE = 0.31, RMSE = 0.43, MAD = 0.8) and a satisfactory level of explanatory accuracy ($R^2 = 0.72$). These findings indicate that the proposed system can deliver stable and meaningful forecasts, thereby serving as a practical decision-support instrument for academic performance evaluation and research planning within higher education institutions.

Bibliography

[1] P. Amaliyah, P. F. Aulia, M. F. Akbar, R. A. Maulidana, and A. Safitri, "Analisis dan desain sistem aplikasi penjadwalan kuliah pengganti (Reschedule-In) bagi mahasiswa dan dosen di universitas," JDBIM (Journal of Digital Business and Innovation Management), vol. 2, no. 1, pp. 34–48, 2023.

[2] M. M. Aminy, "Bibliometric analysis of publications by top 10 Indonesian universities ranked by QS World University Rankings 2022 indexed in PubMed.gov database," International Journal of Health Sciences, vol. 6, no. S9, pp. 294–302, 2022.

[3] Y. Aryani, I. Aqil, and B. Paramita, "Penerapan unified modeling language (UML) pada digitalisasi sistem informasi perpustakaan," Digital Transformation Technology (Digitech), vol. 4, no. 2, pp. 1032–1040, 2025.

[4] Z. Elhachemi, H. Bourouba, and S. Bourouba, "Automatic complexity analysis of UML class diagrams using visual question answering (VQA) techniques," Software, vol. 4, p. 22, 2025.

[5] P. Górski and W. Stecz, "A method for modeling and testing near-real-time system scenarios," Applied Sciences, vol. 14, no. 4, p. 2023, 2024.

[6] Z. Liu, Z. Zhang, and W. Zhang, "A hybrid framework integrating traditional models and deep learning for multi-scale time series forecasting," Entropy, vol. 27, no. 7, p. 695, 2025.

[7] Rahman, "Analisis bibliometrik perkembangan penelitian: Studi literatur sistematis dengan menggunakan VOSviewer," Matra Pembaruan, vol. 7, no. 1, pp. 37–48, 2023.

[8] Rasiban, A. Septiansyah, S. Hasanah, V. N. Permatasari, and A. Yuliawati, "Sistem informasi otomatisasi pelaporan data penjualan toko buku Nazwa yang masuk dan buku yang keluar," IKRAITH-INFORMATIKA, vol. 8, no. 1, pp. 279–292, 2024.

[9] A. Sudewo, J. A. Aritonang, and S. Putra, "Analisis prediksi harapan lama sekolah di Sumatera Utara menggunakan metode ARIMA," Sistemasi: Jurnal Sistem Informasi, vol. 13, no. 6, pp. 2509–2518, 2024.

[10] M. Yıldız and T. Karakuş Yılmaz, "Bibliometric analysis in scientific research using R: A review of Scopus and Web of Science databases," Journal of Data Applications, vol. 2, pp. 31–46, 2023.

[11] B. Alturas, “Connection between UML use case diagrams and UML class diagrams: A matrix proposal,” International Journal of Computer Applications in Technology, vol. 72, no. 3, pp. 161–168, 2023.
[12] A. Casolaro, V. Capone, G. Iannuzzo, and F. Camastra, “Deep learning for time series forecasting: Advances and open problems,” Information, vol. 14, p. 598, 2023.
[13] P. Istiana and P. Wicaksono, “Bibliometric analysis of Indonesian Journal of Geography from 2015–2022,” Indonesian Journal of Geography, vol. 57, no. 1, pp. 80–85, 2025.
[14] H. F. Moed, Citation Analysis in Research Evaluation. Cham, Switzerland: Springer, 2022.
[15] J. Torcal, V. Moreno, J. Llorens, and A. Granados, “Creating and validating a ground truth dataset of Unified Modeling Language diagrams using deep learning techniques,” Applied Sciences, vol. 14, p. 10873, 2024.
[16] S. Sunardi, “Penerapan algoritma K-Nearest Neighbors (KNN) berbasis Django sebagai sistem pendukung keputusan dalam pemilihan calon anggota BEM,” TRILOGI: Jurnal Ilmu Teknologi, Kesehatan, dan Humaniora, vol. 6, no. 1, pp. 110–118, 2025.
[17] C. Andreas, “Perbandingan pendekatan statistika dan machine learning dalam peramalan data time series,” Jurnal Sistem Informasi TGD, vol. 4, no. 4, pp. 778–785, 2025.
[18] M. Silva and C. T. Mesquita, “Publication of a scientific article: What authors need to know,” International Journal of Cardiovascular Sciences, vol. 36, 2023.
[19] M. Almalki, “AI-driven decision support systems in agile software project management: Enhancing risk mitigation and resource allocation,” Systems, vol. 13, no. 4, p. 208, 2025.
[20] B. Ferreira, S. Marques, M. Kalinowski, H. Lopes, and S. D. J. Barbosa, “Lessons learned to improve the UX practices in agile projects involving data science and process automation,” Information and Software Technology, vol. 141, p. 106724, 2022.
[21] M. Hanslo and M. Tanner, “Enhancing agile methodology adoption with machine learning models,” in Proc. Federated Conf. on Computer Science and Information Systems (FedCSIS), vol. 21, pp. 697–704, 2022.
[22] O. D. Simpson, J. Timasni, I. Gunawan, H. Mardiana, and R. Setiawan, “Leveraging big data to accelerate agile transformation in technology firms,” Engineering Science & Technology Journal, vol. 5, no. 6, pp. 1952–1968, 2024.