UNIVERSITAS
GADJAH MADA
Implementasi Time Series Forecasting Berbasis Machine Learning untuk Autoscaling Proaktif Pada Kubernetes
Safarini Gandira, Ir. Yuris Mulya Saputra, S.T., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

[1] J. Lourenço and A. R. Silva, "Monolith Development History for Microservices Identification: a Comparative Analysis," in *2023 IEEE International Conference on Web Services (ICWS)*, Chicago, IL, USA: IEEE, Jul. 2023, pp. 50–56. doi: 10.1109/ICWS60048.2023.00019.

[2] M. M. Raikar *et al.*, "Leveraging Docker Containers for Deployment of Web Applications in Microservices Architecture," in *2024 First International Conference for Women in Computing (InCoWoCo)*, Pune, India: IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/InCoWoCo64194.2024.10863683.

[3] F. B. Fava *et al.*, "Assessing the Performance of Docker in Docker Containers for Microservice-Based Architectures," in *2024 32nd Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, Dublin, Ireland: IEEE, Mar. 2024, pp. 137–142. doi: 10.1109/PDP62718.2024.00026.

[4] "Kubernetes Documentation." Accessed: May 26, 2026. [Online]. Available: https://kubernetes.io/docs/concepts/workloads/autoscaling/

[5] V. Pandey, "Cost-Optimized Predictive Autoscaling of Cloud Resources Using Game-Theoretic Queuing Theory," in *SoutheastCon 2025*, Concord, NC, USA: IEEE, Mar. 2025, pp. 481–487. doi: 10.1109/SoutheastCon56624.2025.10971700.

[6] I. K. Kim, W. Wang, Y. Qi, and M. Humphrey, "Forecasting Cloud Application Workloads With *CloudInsight* for Predictive Resource Management," *IEEE Trans. Cloud Comput.*, vol. 10, no. 3, pp. 1848–1863, Jul. 2022, doi: 10.1109/TCC.2020.2998017.

[7] M. P. J. Kuranage, E. Hanser, A. Bouabdallah, L. Nuaymi, and P. Bertin, "CPU throttling-aware AI-based autoscaling for Kubernetes," in *2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Valencia, Spain: IEEE, Sep. 2024, pp. 1–7. doi: 10.1109/PIMRC59610.2024.10817283.

[8] D. G. Taslim and I. M. Murwantara, "A Comparative Study of ARIMA and LSTM in Forecasting Time Series Data," in *2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Semarang, Indonesia: IEEE, Aug. 2022, pp. 231–235. doi: 10.1109/ICITACEE55701.2022.9924148.

[9] O. Zarai, Z. Mcharfi, and B. El Asri, "Intelligent Autoscaling Strategies for Cloud-Native Microservices: A Systematic Review," in *2025 International Conference on Intelligent Systems: Theories and Applications (SITA)*, Rabat, Morocco: IEEE, Oct. 2025, pp. 1–9. doi: 10.1109/SITA67914.2025.11273589.

[10] S. Kakade *et al.*, "Proactive Horizontal Pod Autoscaling in Kubernetes using Bi-LSTM," in *2023 IEEE International Conference on Contemporary Computing and Communications (InC4)*, Bangalore, India: IEEE, Apr. 2023, pp. 1–5. doi: 10.1109/InC457730.2023.10263031.

UNIVERSITAS
GADJAH MADA
Implementasi Time Series Forecasting Berbasis Machine Learning untuk Autoscaling Proaktif Pada Kubernetes
Safarini Gandira, Ir. Yuris Mulya Saputra, S.T., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

[11] N.-M. Dang-Quang and M. Yoo, "Deep Learning-Based Autoscaling Using Bidirectional Long Short-Term Memory for Kubernetes," *Appl. Sci.*, vol. 11, no. 9, p. 3835, Apr. 2021, doi: 10.3390/app11093835.

[12] J. Xiao, F. Yang, W. Ai, T. Meng, and G. Xie, "PAHPA: Revolutionizing Kubernetes Autoscaling with Integrated Predictive Analytics and Real-Time Monitoring," *IEEE Trans. Serv. Comput.*, pp. 1–13, 2026, doi: 10.1109/TSC.2026.3685325.

[13] A. A. Pramesti and A. I. Kistijantoro, "Autoscaling Based on Response Time Prediction for Microservice Application in Kubernetes," in *2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, Tokoname, Japan: IEEE, Sep. 2022, pp. 1–6. doi: 10.1109/ICAICTA56449.2022.9932943.

[14] M. Yan, X. Liang, Z. Lu, J. Wu, and W. Zhang, "HANSEL: Adaptive horizontal scaling of microservices using Bi-LSTM," *Appl. Soft Comput.*, vol. 105, p. 107216, Jul. 2021, doi: 10.1016/j.asoc.2021.107216.

[15] P. B. Guruge and Y. H. P. P. Priyadarshana, "Time series forecasting-based Kubernetes autoscaling using Facebook Prophet and Long Short-Term Memory," *Front. Comput. Sci.*, vol. 7, p. 1509165, Feb. 2025, doi: 10.3389/fcomp.2025.1509165.

[16] Sahala Wahyu Wardana, "Adaptif Horizontal Autoscaling Pada Platform Kubernetes Dengan Bidirectional Long-Short Term Memory (Bi-LSTM) Dan Graph Neural Network (GNN)," Universitas Gadjah Mada, 2025.

[17] H. Beshley, S. Vilhutskyi, M. Beshley, and M. Greguš, "AI-Based Proactive Autoscaling Technique for Web Applications in Kubernetes," in *2026 IEEE 18th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv, Ukraine: IEEE, Feb. 2026, pp. 1–6. doi: 10.1109/TCSET65181.2026.11461141.

[18] A. A. Khaleq and I. Ra, "Intelligent Autoscaling of Microservices in the Cloud for Real-Time Applications," *IEEE Access*, vol. 9, pp. 35464–35476, 2021, doi: 10.1109/ACCESS.2021.3061890.

[19] N. Shafi, M. Abdullah, W. Iqbal, A. Erradi, and F. Bukhari, "Cdascaler: a cost-effective dynamic autoscaling approach for containerized microservices," *Clust. Comput.*, vol. 27, no. 4, pp. 5195–5215, Jul. 2024, doi: 10.1007/s10586-023-04228-y.

[20] L. De Lauretis, "From Monolithic Architecture to Microservices Architecture," in *2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, Berlin, Germany: IEEE, Oct. 2019, pp. 93–96. doi: 10.1109/ISSREW.2019.00050.

[21] Oyekunle Claudius Oyeniran, Adebunmi Okechukwu Adewusi, Adams Gbolahan Adeleke, Lucy Anthony Akwawa, and Chidimma Francisca Azubuko, "Microservices architecture in cloud-native applications: Design patterns and scalability," *Comput. Sci. IT Res. J.*, vol. 5, no. 9, pp. 2107–2124, Sep. 2024, doi: 10.51594/csitrj.v5i9.1554.

[22] D. Shi, J. Zhang, Y. Deng, J. Pan, X. Zheng, and J. Qian, "Containerization Technology for Real-Time Data Collaborative Processing in Power Dispatch Systems," in *2025 IEEE 12th Joint International Information Technology and*

UNIVERSITAS
GADJAH MADA
Implementasi Time Series Forecasting Berbasis Machine Learning untuk Autoscaling Proaktif Pada Kubernetes
Safarini Gandira, Ir. Yuris Mulya Saputra, S.T., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Artificial Intelligence Conference (ITAIC), Chongqing, China: IEEE, May 2025, pp. 392–396. doi: 10.1109/ITAIC64559.2025.11162977.

[23] S. Liu and P. Lv, "Short-term Load Forecasting Model Based on Attention Mechanism and Gated Recurrent Unit," in *2019 IEEE 8th International Conference on Advanced Power System Automation and Protection (APAP)*, Xi'an, China: IEEE, Oct. 2019, pp. 1784–1788. doi: 10.1109/APAP47170.2019.9225191.

[24] A. Bhardwaj, K. Pandey, R. Sharma, L. Deewan, and U. Gupta, "Comparative Performance Analysis of Docker Container and Virtual Machine: A Comprehensive Evaluation," in *2025 International Conference on Next Generation Information System Engineering (NGISE)*, Ghaziabad, Delhi (NCR), India: IEEE, Mar. 2025, pp. 1–6. doi: 10.1109/NGISE64126.2025.11085383.

[25] Z. Ding and Q. Huang, "COPA: A Combined Autoscaling Method for Kubernetes," in *2021 IEEE International Conference on Web Services (ICWS)*, Chicago, IL, USA: IEEE, Sep. 2021, pp. 416–425. doi: 10.1109/ICWS53863.2021.00061.

[26] M.-N. Tran, D.-D. Vu, and Y. Kim, "A Survey of Autoscaling in Kubernetes," in *2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN)*, Barcelona, Spain: IEEE, Jul. 2022, pp. 263–265. doi: 10.1109/ICUFN55119.2022.9829572.

[27] S. Prakash, A. S. Jalal, and P. Pathak, "Forecasting COVID-19 Pandemic using Prophet, LSTM, hybrid GRU-LSTM, CNN-LSTM, Bi-LSTM and Stacked-LSTM for India," in *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India: IEEE, Mar. 2023, pp. 1–6. doi: 10.1109/ISCON57294.2023.10112065.

[28] B. Hrnjica and O. Bonacci, "Lake Level Prediction using Feed Forward and Recurrent Neural Networks," *Water Resour. Manag.*, vol. 33, no. 7, pp. 2471–2484, May 2019, doi: 10.1007/s11269-019-02255-2.

[29] F. H. Putti, M. A. Murti, A. Novianty, C. Setianingsih, and N. H. Ismail, "Social Media Analysis Using LSTM and Bi-LSTM for Automated Identification of Depression Cancer Patients," in *2025 10th International Conference on Computer and Communication Engineering (ICCCE)*, Kuala Lumpur, Malaysia: IEEE, Aug. 2025, pp. 105–110. doi: 10.1109/ICCCE66530.2025.11474067.

[30] S. Cornegruta, R. Bakewell, S. Withey, and G. Montana, "Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks," Sep. 27, 2016, *arXiv*: arXiv:1609.08409. doi: 10.48550/arXiv.1609.08409.

[31] Z. Zhou and T. Zhan, "The Improved Bi-LSTM-Transformer Model and its Application," in *2024 5th International Conference on Computer, Big Data and Artificial Intelligence (ICCBD+AI)*, Jingdezhen, China: IEEE, Nov. 2024, pp. 71–75. doi: 10.1109/ICCBD-AI65562.2024.00019.

[32] M. Robith, R. S. Wibowo, and W. Wibowo, "Short-Term Wind Power Forecasting in East Java Using Gated Recurrent Unit," in *2023 3rd International Conference on Electronic and Electrical Engineering and*

UNIVERSITAS
GADJAH MADA
Implementasi Time Series Forecasting Berbasis Machine Learning untuk Autoscaling Proaktif Pada Kubernetes
Safarini Gandira, Ir. Yuris Mulya Saputra, S.T., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Intelligent System (ICE3IS), Yogyakarta, Indonesia: IEEE, Aug. 2023, pp. 58–63. doi: 10.1109/ICE3IS59323.2023.10335218.

[33] B. Arundhati, Y. Agrapu, L. Gandipadala, B. Gundu, M. Chokkakula, and H. H. Karrimajji, "Wind Energy Prediction Using Machine Learning Techniques," in *2026 4th International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, CHICKABALLAPURA, India: IEEE, Apr. 2026, pp. 1–5. doi: 10.1109/ICKECS70176.2026.11527871.

[34] "Microservices Demo." [Online]. Available: https://github.com/googlecloudplatform/microservices-demo

[35] Roman Arcea, "The surprising economics of Horizontal Pod Autoscaling tuning." [Online]. Available: https://cloud.google.com/blog/products/containers-kubernetes/tuning-the-kubernetes-hpa-in-gke