

Model Prediksi Data Time Series Menggunakan KAN (Kolmogorov Arnold Network)

Arwansyah^{1*}, Ahyuna², Hasryif SY³, Suryani⁴, Nurdiansah⁵, Samsu Alam⁶

^{1,2,3,4,5,6}Universitas Dipa Makassar; Jl. Perintis Kemerdekaan Km. 9 Makassar, (0411)587194

^{1,2,3,4,6}Teknik Informatika, ⁵Sistem Informasi, Universitas Dipa Makassar

e-mail: ¹arwansyah@undipa.ac.id, ²ahyuna@undipa.ac.id, ³hasyirif@undipa.ac.id,

⁴suryani187@undipa.ac.id, ⁵nurdiansah@undipa.ac.id, ⁶alam@undipa.ac.id

Abstrak

Studi ini memperkenalkan dan mengevaluasi Kolmogorov-Arnold Network (KAN) sebagai solusi yang efisien dan dapat diinterpretasikan untuk peramalan deret waktu, yang bertujuan untuk mengatasi kompleksitas dan ketidakjelasan model seperti LSTM dan Transformer. Metodologi penelitian melibatkan pengujian model KAN pada tugas prediksi jangka panjang menggunakan dataset ETTh1, ETTm1, dan Nilai Tukar, dengan panjang prediksi yang mencakup interval 3, 6, 9, 12, dan 24. Dataset, yang mencakup frekuensi per jam, menit, dan harian, diproses terlebih dahulu secara univariat melalui normalisasi MinMaxScaler dan format jendela geser. Sebagai pendekatan baru yang dibangun di atas representasi fungsi universal, arsitektur KAN menggunakan aktivasi non-linier seperti fungsi sinus untuk menangkap dinamika temporal yang kompleks. Dengan menggunakan Mean Squared Error (MSE) dan Mean Absolute Error (MAE) sebagai metrik evaluasi, eksperimen kami menghasilkan hasil yang tepat secara konsisten di semua skenario. Skor metrik rata-rata adalah 2,129 untuk ETTh1, 0,670 untuk ETTm1, dan 0,025 untuk dataset nilai tukar. Kemampuan arsitektur KAN yang telah terbukti untuk memodelkan dependensi berbasis waktu menegaskan potensinya sebagai alternatif yang kompetitif dalam bidang analisis deret waktu.

Kata kunci: Time Series, Prediksi, Kolmogorov-Arnold Network, Forecasting, Univariate.

Abstract

This study introduces and evaluates the Kolmogorov-Arnold Network (KAN) as an efficient and interpretable solution for time series forecasting, aiming to overcome the complexity and opacity of models like LSTM and Transformer. The research methodology involved testing the KAN model on long-term prediction tasks using the ETTh1, ETTm1, and Exchange Rate datasets, with prediction lengths spanning 3, 6, 9, 12, and 24 intervals. The datasets, which include hourly, minute, and daily frequencies, were pre-processed univariately through MinMaxScaler normalization and a sliding window format. As a novel approach built on universal function representation, the KAN architecture employs non-linear activations like sine functions to capture complex temporal dynamics. Using Mean Squared Error (MSE) and Mean Absolute Error (MAE) as evaluation metrics, our experiments yielded consistently precise results across all scenarios. The average metric scores were 2.129 for ETTh1, 0.670 for ETTm1, and 0.025 for the exchange rate dataset. The demonstrated ability of the KAN architecture to model time-based dependencies confirms its potential as a competitive alternative in the field of time series analysis.

Keywords: Time Series, Prediction, Kolmogorov-Arnold Network, Forecasting, Univariate.

1. PENDAHULUAN

Prediksi data deret waktu (time series) merupakan salah satu pilar utama dalam pengambilan keputusan berbasis data di berbagai sektor krusial. Akurasi peramalan sangat vital dalam bidang-bidang seperti perencanaan energi, peramalan cuaca, dan terutama manajemen risiko keuangan, di mana keputusan strategis bergantung pada proyeksi masa depan (De Gooijer & Hyndman, 2006; Wen et al., 2017). Seiring dengan ledakan volume data, kebutuhan akan model prediksi yang tidak hanya akurat tetapi juga mampu menangani ketidakpastian, seperti dalam peramalan probabilistik, menjadi semakin mendesak (Salinas et al., 2020). Selama bertahun-tahun, metode statistik klasik seperti ARIMA (Autoregressive Integrated Moving Average) telah menjadi standar (Box et al., 2015), namun seringkali kesulitan menangkap pola non-linier yang kompleks pada data dunia nyata (Hewamalage et al., 2021).

Perkembangan pesat dalam bidang deep learning telah menghadirkan solusi yang lebih kuat. Kemunculan arsitektur Recurrent Neural Network (RNN) dan variannya menjadi titik balik. Arsitektur Long Short-Term Memory (LSTM) oleh Hochreiter & Schmidhuber (1997) secara khusus merevolusi kemampuan model untuk mempelajari dependensi jangka panjang. Selain LSTM, varian lain seperti Gated Recurrent Unit (GRU) juga menunjukkan performa kompetitif dengan arsitektur yang lebih sederhana (Cho et al., 2014). Inovasi lebih lanjut melahirkan model hibrida seperti LSTNet, yang menggabungkan CNN dan RNN (Lai et al., 2018), serta arsitektur non-rekuren seperti WaveNet yang menggunakan dilated convolutions dan terbukti sangat efektif untuk data sekuensial (Oord et al., 2016).

Era terkini didominasi oleh kesuksesan arsitektur Transformer yang diperkenalkan oleh Vaswani et al. (2017). Meskipun awalnya dirancang untuk pemrosesan bahasa, prinsip self-attention yang digunakannya terbukti sangat efektif untuk tugas prediksi deret waktu. Hal ini memicu lahirnya model-model canggih seperti Informer (Zhou et al., 2021), Autoformer (Wu et al., 2021), dan PatchTST (Nie et al., 2023), yang menunjukkan kinerja terdepan dalam skenario prediksi jangka panjang. Model-model inilah yang juga mempopulerkan penggunaan dataset benchmark seperti ETTh, ETTm, dan Exchange Rate yang digunakan dalam penelitian ini.

Meskipun model berbasis LSTM dan Transformer menunjukkan akurasi yang impresif, mereka tidak lepas dari tantangan signifikan. Salah satu kendala utama adalah kompleksitas model yang sangat tinggi dan kebutuhan sumber daya komputasi yang besar, terutama pada model Transformer dengan mekanisme perhatiannya yang berskala kuadratik (Kitaev et al., 2020). Tantangan kedua adalah sifat "kotak hitam" (black box) yang melekat pada model-model ini (Sezer et al., 2020). Sulit untuk menginterpretasikan dasar pengambilan keputusan model, sebuah kekurangan fatal dalam domain yang menuntut transparansi dan akuntabilitas (Ribeiro et al., 2016). Kedua tantangan ini menciptakan sebuah celah penelitian: kebutuhan akan arsitektur model yang tidak hanya akurat, tetapi juga efisien dan dapat diinterpretasikan (Lim & Zohren, 2021).

Studi ini memperkenalkan dan mengevaluasi Kolmogorov-Arnold Network (KAN), yang baru-baru ini diusulkan oleh Liu et al. (2024), sebagai solusi yang efisien dan dapat diinterpretasikan untuk peramalan deret waktu. KAN secara fundamental berbeda dari arsitektur standar karena menempatkan fungsi aktivasi yang dapat dipelajari (learnable) pada setiap sisi (edge) jaringan, bukan fungsi aktivasi tetap pada simpul (node). Pendekatan ini terinspirasi langsung dari Teorema Representasi Kolmogorov-Arnold (Kolmogorov, 1957) dan berpotensi mencapai akurasi lebih tinggi dengan parameter yang jauh lebih sedikit. Lebih penting lagi, struktur KAN memungkinkan visualisasi, menawarkan tingkat interpretasi yang tidak dapat dicapai oleh model konvensional. Dengan proses pelatihan yang dapat dioptimalkan menggunakan metode standar seperti Adam (Kingma & Ba, 2014), KAN diposisikan sebagai alternatif yang menjanjikan. Oleh karena itu, penelitian ini akan menginvestigasi penerapan KAN sebagai model prediksi deret waktu yang ringan, akurat, dan efisien.

Penelitian ini bertujuan untuk menginvestigasi secara mendalam performa dan efektivitas model Kolmogorov-Arnold Network (KAN) dalam tugas prediksi data time series. Untuk menjaga fokus, penelitian ini membatasi ruang lingkupnya pada data time series univariat. Performa KAN akan diuji menggunakan tiga jenis dataset utama: energi (ETTh1 dan ETTm1) serta keuangan (Exchange Rate). Dataset ETTh1 dan ETTm1 merepresentasikan data beban listrik pada skala jam dan menit, sedangkan Exchange Rate mencerminkan dinamika pasar keuangan. Ketiga dataset ini dipilih karena dianggap cukup untuk mewakili karakteristik data yang beragam dari domain yang berbeda, dan akan dievaluasi menggunakan skenario panjang prediksi yang bervariasi (3, 6, 9, 12, dan 24 langkah ke depan). Kami juga membandingkan kinerja KAN dengan model lain pada literatur. Sebagai contoh, Evaluasi ini dilakukan untuk menjawab apakah KAN dapat menjadi alternatif model yang lebih ringan dan efisien secara komputasi dibandingkan arsitektur kompleks seperti LSTM dan Transformer. Diharapkan penelitian ini dapat memberikan kontribusi akademis pada kajian penerapan KAN di ranah deret waktu, sekaligus menawarkan solusi praktis yang akurat dan hemat sumber daya bagi sistem dengan keterbatasan komputasi.

2. METODE PENELITIAN

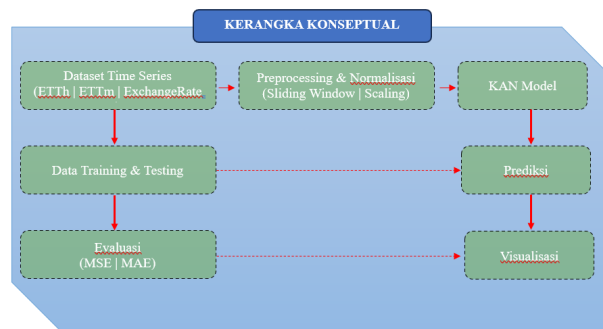
Penelitian ini disusun untuk mengatasi beberapa permasalahan dan mencapai tujuan penelitian dengan pendekatan eksperimental menggunakan model Kolmogorov-Arnold Network (KAN) pada data time series univariat.

2.1 Kerangka Konseptual

Kerangka konseptual dalam penelitian ini menggambarkan alur proses penggunaan model Kolmogorov-Arnold Network (KAN) dalam memprediksi data time series. KAN digunakan sebagai pendekatan alternatif terhadap model deep learning konvensional karena kemampuannya dalam melakukan pendekatan fungsi nonlinier secara efisien dan ringan. Penelitian ini juga melibatkan evaluasi

kinerja KAN terhadap beberapa dataset dari domain berbeda dengan berbagai panjang horizon prediksi (*prediction length*).

1. Dataset Time Series: Dataset berupa data historis yang direkam secara periodik, seperti ETTh1, ETTm1, dan Exchange Rate.
2. Preprocessing & Normalisasi: Tahapan ini mencakup pemilihan variabel target, pembentukan *sliding window*, dan normalisasi data menggunakan MinMaxScaler.
3. Model Kolmogorov-Arnold Network (KAN): Model ini terdiri dari dua lapisan linear dengan aktivasi sinusoidal di antaranya, yang mampu mempelajari pola nonlinier dari data.
4. Evaluasi Model: Melibatkan pengukuran performa model menggunakan metrik MSE dan MAE untuk menilai akurasi prediksi pada berbagai *prediction length*.
5. Hasil dan Visualisasi: Hasil berupa nilai prediksi dan aktual, ditampilkan dalam bentuk visualisasi grafik serta tabel yang menggambarkan performa model secara kuantitatif.



Gambar 1. Kerangka Konseptual

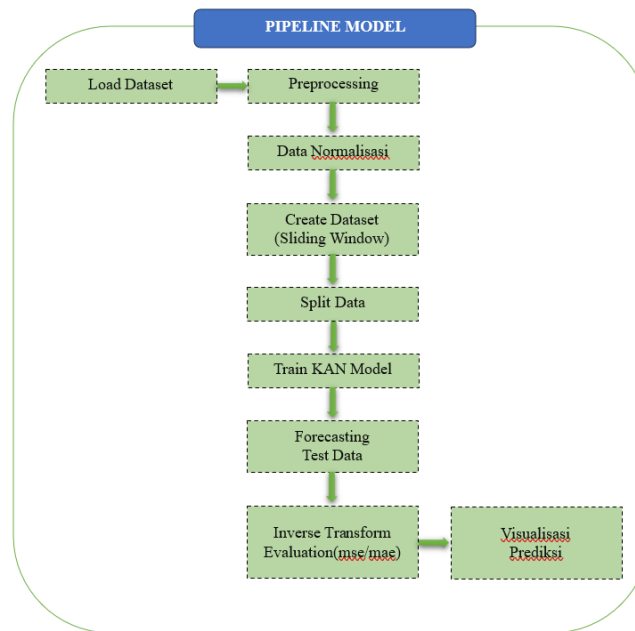
2.2. Dataset

Penelitian ini menggunakan empat dataset yang relevan untuk mengevaluasi model time series berbasis Kolmogorov-Arnold Network (KAN). Dataset yang dipilih adalah ETTh, ETTm, dan Exchange Rate diunduh dari repositori open-source. Data difokuskan pada satu variabel target (univariate). Dataset dibagi dalam format (X, y) menggunakan pendekatan sliding window misalnya *lookback* = 24 dan *prediction length* (3, 6, 9, 12, 24).

2.3. Pra-Pemrosesan Data

Sebelum data digunakan untuk pelatihan model, beberapa langkah pra-pemrosesan dilakukan:

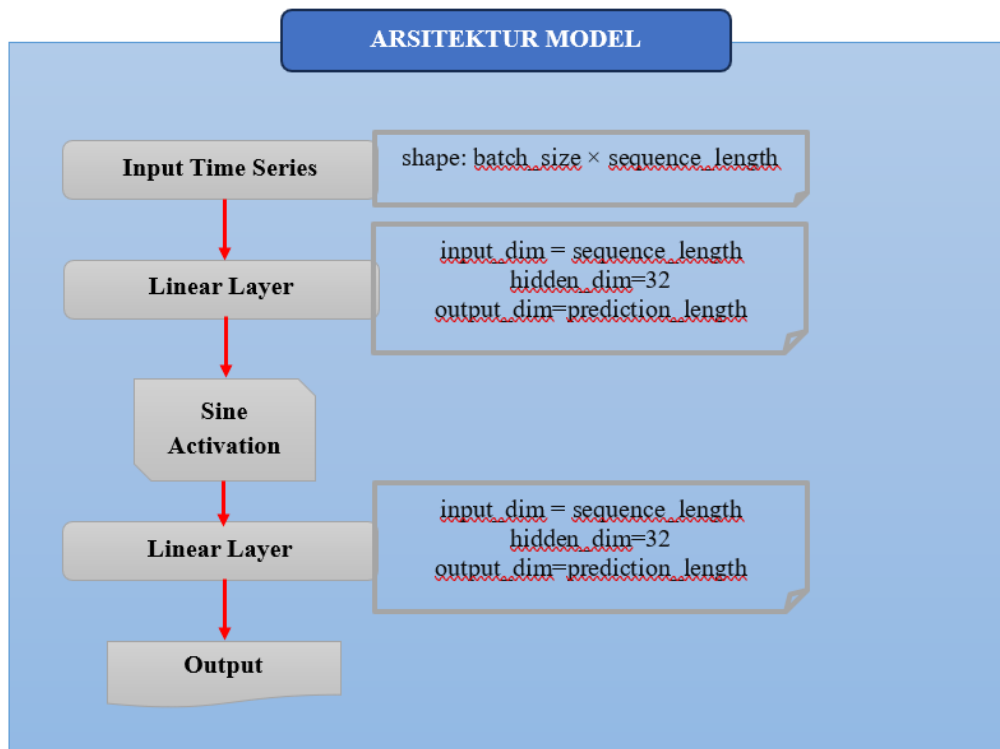
1. Pembersihan Data: Mengidentifikasi dan menangani data yang hilang, duplikat, atau tidak konsisten. Langkah ini memastikan bahwa data yang digunakan dalam model bersih dan siap untuk analisis.
2. Normalisasi dan Standarisasi: Menormalkan atau menstandarisasi data untuk memastikan bahwa semua fitur memiliki skala yang sama. Ini membantu dalam mempercepat proses pelatihan dan meningkatkan kinerja model.
3. Transformasi Data: Mengubah data menjadi format yang sesuai untuk model deep learning.



Gambar 2. Pipeline Model

2.4. Arsitektur Model

Penelitian ini menggunakan pendekatan kuantitatif eksperimental, di mana dilakukan serangkaian pengujian model KAN terhadap tiga dataset time series yang berbeda. Hasil prediksi dianalisis menggunakan metrik evaluasi untuk menentukan performa model pada berbagai *prediction length*. Pada model KAN, representasi non-linear tidak berasal dari aktivasi eksplisit seperti ReLU, melainkan melalui fungsi spline atau fungsi sinus yang menyerap kompleksitas fungsi target secara efisien.



Gambar 3. Arsitektur Model

2.4. Hyperparameter

Beberapa hyperparameter yang digunakan meliputi:

1. Sequence Length: 12

2. Prediction Length: 3,6,9,12,24
3. Loss Function: MSELoss
4. Optimizer: ADAM
5. Batch size: 32
6. Learning rate: 0.001
7. Epoch: 50
8. Jumlah neuron: 32

3. HASIL DAN PEMBAHASAN

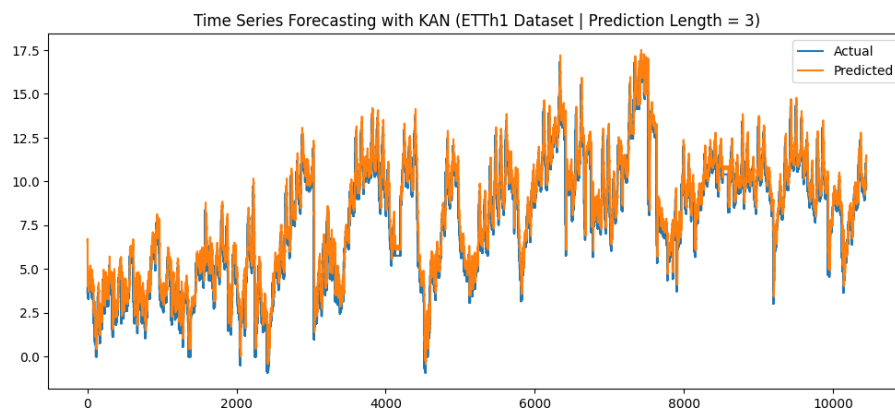
Hasil dan pembahasan dalam penelitian ini berfokus pada implementasi model Kolmogorov-Arnold Network (KAN) pada data time series univariate. Hasil experiment dapat dilihat pada table berikut:

Tabel 1. Matriks Hasil Prediksi

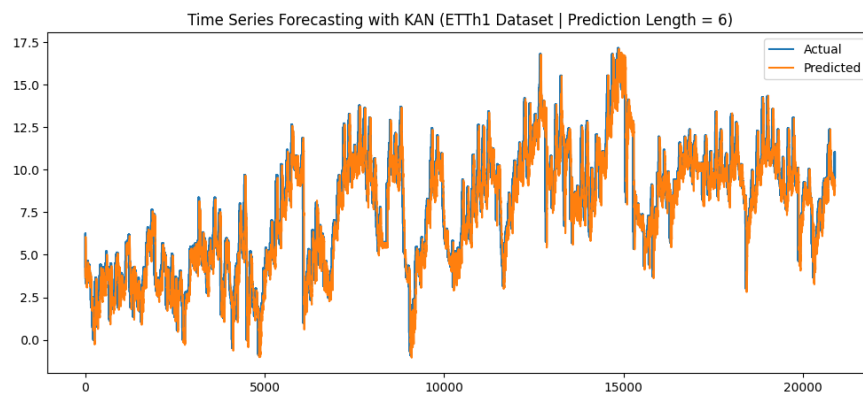
Dataset	Prediction Length									
	3		6		9		12		24	
	mse/mae		mse/mae		mse/mae		mse/mae		mse/mae	
ETTh1	[1.031][0.738]		[1.621][0.901]		[2.062][1.027]		[2.728][1.204]		[3.203][1.308]	
ETTm1	[0.236][0.331]		[0.394][0.415]		[0.608][0.529]		[0.727][0.564]		[1.387][0.807]	
ExchangeRate	[0.001][0.006]		[0.001][0.006]		[0.001][0.007]		[0.002][0.009]		[0.012][0.002]	

3.1 Grafik Prediksi

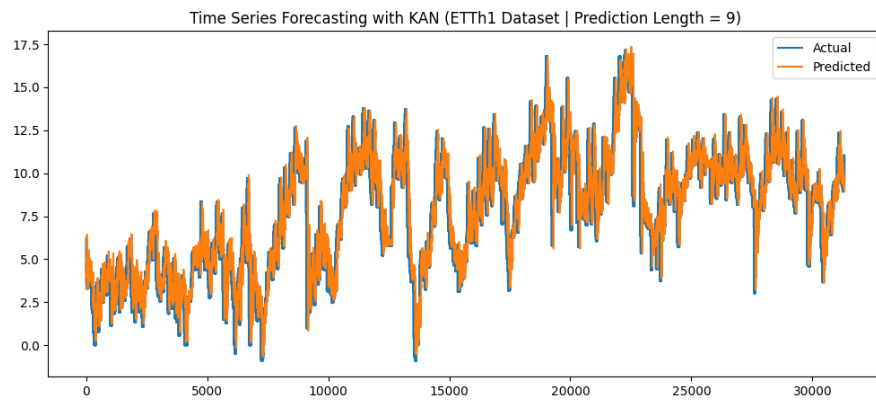
1. ETTh1 Dataset



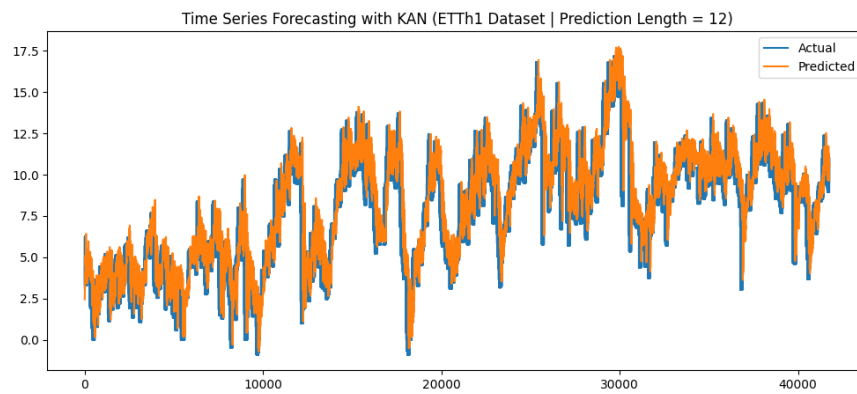
Gambar 4. Grafik Actual vs Predicted Prediction Length = 3



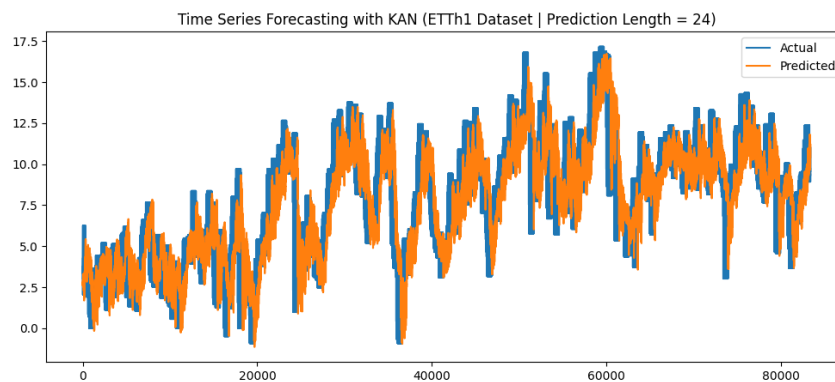
Gambar 5. Grafik Actual vs Predicted Prediction Length = 6



Gambar 6. Grafik Actual vs Predicted Prediction Length = 9

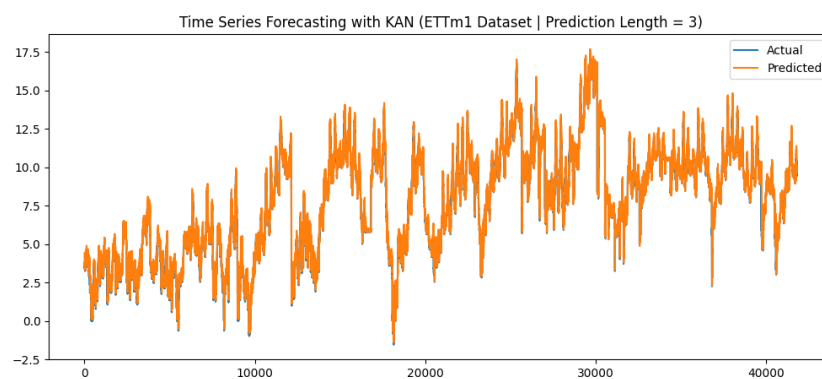


Gambar 7. Grafik Actual vs Predicted Prediction Length = 12

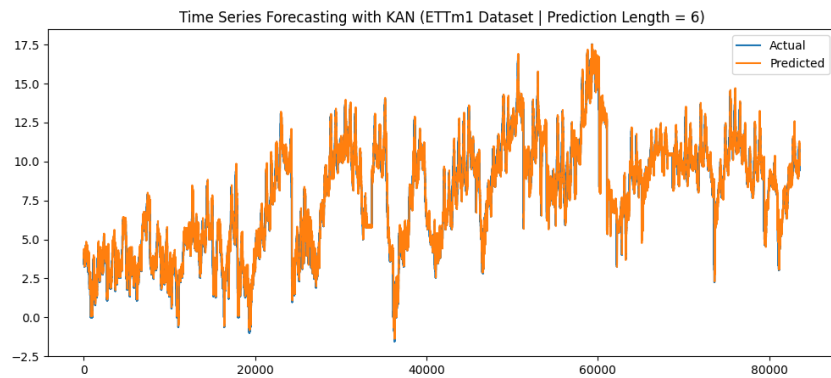


Gambar 8. Grafik Actual vs Predicted Prediction Length = 24

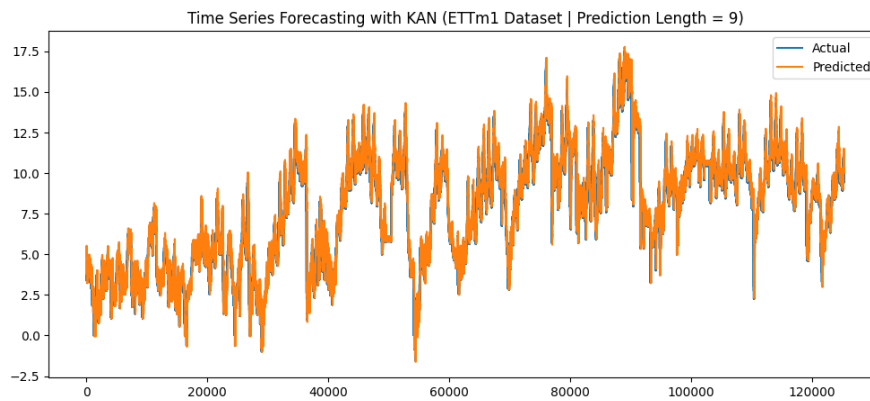
2. *ETTh1* Dataset



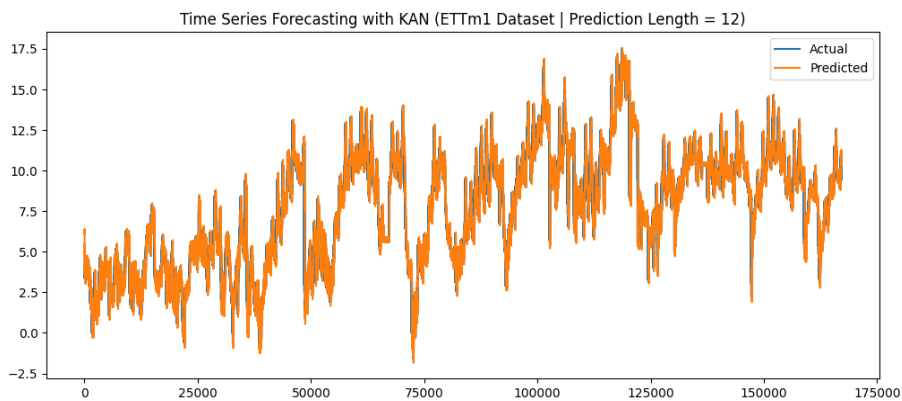
Gambar 9. Grafik Actual vs Predicted Prediction Length = 3



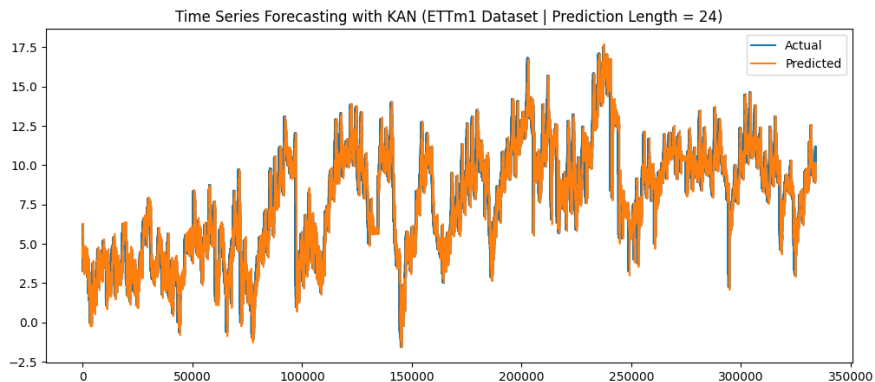
Gambar 10. Grafik Actual vs Predicted Prediction Length = 6



Gambar 11. Grafik Actual vs Predicted Prediction Length = 9

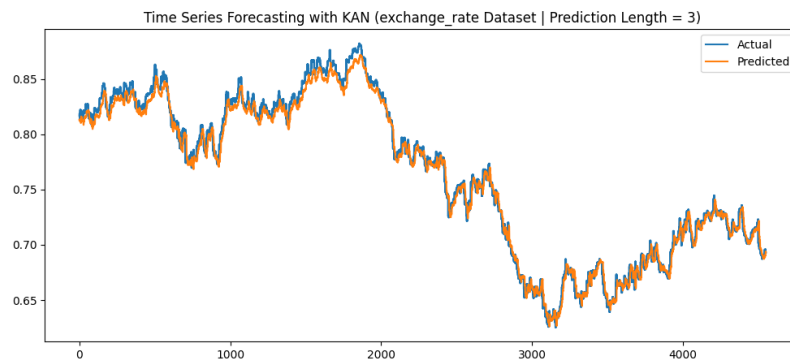


Gambar 12. Grafik Actual vs Predicted Prediction Length = 12

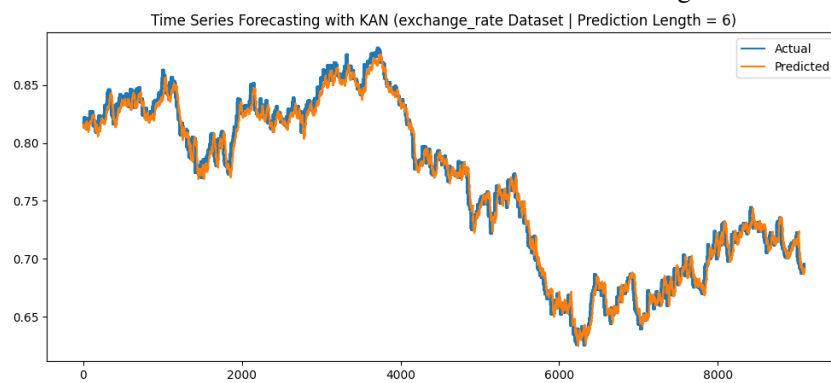


Gambar 13. Grafik Actual vs Predicted Prediction Length = 24

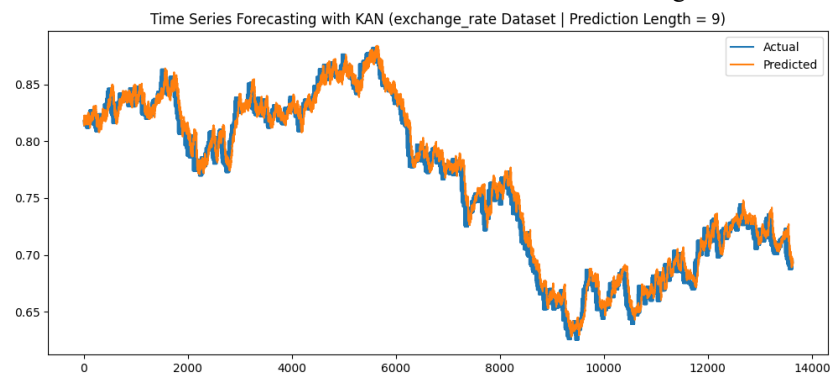
3. Exchange Rate Dataset



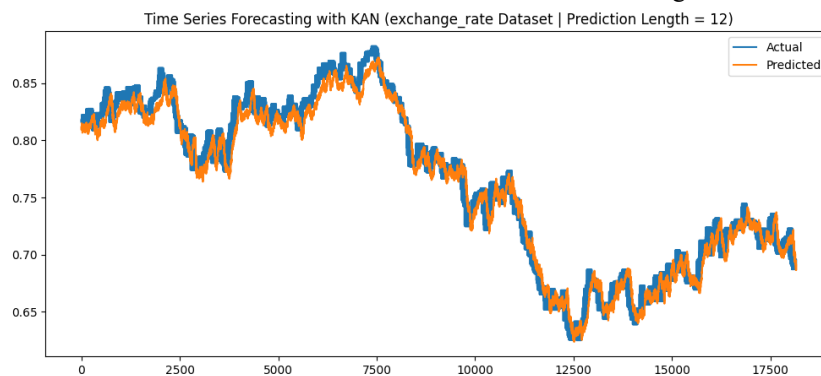
Gambar 14. Grafik Actual vs Predicted Prediction Length = 3



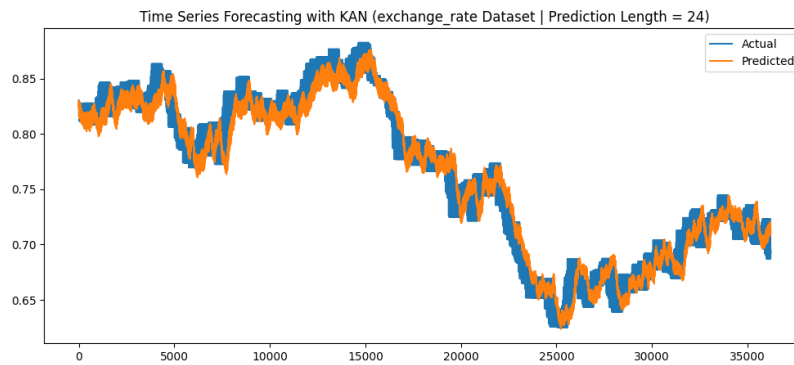
Gambar 15. Grafik Actual vs Predicted Prediction Length = 6



Gambar 16. Grafik Actual vs Predicted Prediction Length = 9



Gambar 17. Grafik Actual vs Predicted Prediction Length = 12



Gambar 18. Grafik Actual vs Predicted Prediction Length = 24

Evaluasi dilakukan menggunakan dua metrik utama, yaitu Mean Squared Error (MSE) dan Mean Absolute Error (MAE) dengan analisa hasil sebagai berikut.

1. ETTh1 Dataset

Pada dataset ETTh1, terlihat bahwa semakin panjang prediksi, nilai MSE dan MAE cenderung meningkat, yang mengindikasikan penurunan akurasi. MSE meningkat dari 1.031 pada horizon 3 menjadi 3.203 pada horizon 24, sedangkan MAE meningkat dari 0.738 menjadi 1.308. Hal ini menunjukkan bahwa model KAN memiliki keterbatasan dalam menangani prediksi jangka panjang pada dataset yang kompleks seperti ETTh1.

2. ETTm1 Dataset

Berbeda dengan ETTh1, model menunjukkan performa yang lebih stabil dan relatif akurat pada ETTm1. Nilai MSE berkisar dari 0.236 hingga 1.387, sementara MAE berkisar dari 0.331 hingga 0.807. Peningkatan kesalahan pada prediksi jangka panjang tetap terjadi, namun tidak sekuat peningkatan yang terjadi pada ETTh1. Hal ini mengindikasikan bahwa model KAN dapat menangkap pola temporal jangka pendek dan menengah dengan cukup baik pada data resolusi menengah (menit).

3. Exchange Rate Dataset

Hasil paling signifikan terlihat pada dataset ExchangeRate, di mana KAN menunjukkan performa sangat baik dengan nilai MSE dan MAE yang sangat kecil dan stabil di seluruh panjang prediksi. MSE berkisar antara **0.001** hingga **0.012**, dan MAE antara **0.006** hingga **0.009**. Ini menunjukkan bahwa model KAN sangat efektif dalam menangani data keuangan yang memiliki fluktuasi halus dan pola linier/kuasilinier.

4. Model KAN bekerja optimal pada dataset dengan pola periodik yang stabil dan kompleksitas rendah, seperti *ExchangeRate*.

5. Untuk dataset kompleks seperti ETTh1, diperlukan tuning lanjutan atau integrasi dengan mekanisme lain (misalnya attention) untuk mempertahankan akurasi jangka panjang.

6. Meskipun KAN merupakan model ringan, hasil menunjukkan keseimbangan yang baik antara akurasi dan efisiensi komputasi untuk sebagian besar dataset, terutama pada prediksi jangka pendek hingga menengah

4. KESIMPULAN

Penelitian ini bertujuan untuk mengevaluasi kinerja Kolmogorov-Arnold Network (KAN) dalam melakukan prediksi data deret waktu (time series) pada berbagai domain data dan horizon prediksi. Berdasarkan hasil eksperimen yang telah dilakukan terhadap tiga dataset benchmark, yaitu ETTh1, ETTm1, dan ExchangeRate, dengan panjang prediksi yang bervariasi (3, 6, 9, 12, dan 24), diperoleh beberapa kesimpulan utama sebagai berikut:

1. KAN menunjukkan performa yang baik pada prediksi jangka pendek dan menengah, khususnya pada dataset dengan pola fluktuasi yang halus seperti ExchangeRate, dengan nilai MSE dan MAE yang sangat kecil dan stabil.
2. Pada dataset yang lebih kompleks seperti ETTh1, performa KAN menurun seiring bertambahnya horizon prediksi, mengindikasikan bahwa kompleksitas pola temporal memerlukan model yang lebih ekspresif atau perluasan arsitektur KAN.
3. Stabilitas performa KAN pada ETTm1 menunjukkan bahwa model ini cukup andal untuk menangani data dengan resolusi waktu menengah dan fluktuasi moderat.
4. KAN dapat dijadikan sebagai alternatif model time series yang ringan dan efisien, serta dapat mengurangi beban komputasi dibandingkan dengan model-model berat seperti LSTM atau Transformer, terutama untuk aplikasi real-time dengan keterbatasan sumber daya.

5. Eksperimen ini membuktikan bahwa KAN layak untuk dikembangkan lebih lanjut, baik dari sisi arsitektur maupun integrasi dengan komponen lain seperti attention mechanism untuk meningkatkan kemampuannya dalam menangani prediksi jangka panjang.

5. SARAN

Berdasarkan temuan dan batasan dalam penelitian ini, terdapat beberapa arah pengembangan yang menarik dan dianjurkan untuk dieksplorasi di masa depan guna memperluas pemahaman dan penerapan Kolmogorov-Arnold Network (KAN) dalam domain prediksi deret waktu:

1. Disarankan untuk mengeksplorasi arsitektur hibrida yang menggabungkan kekuatan KAN dengan model lain. Sebagai contoh, model LSTM-KAN atau GRU-KAN dapat dirancang, di mana lapisan LSTM/GRU berfungsi untuk mengekstraksi dependensi temporal, sementara lapisan KAN digunakan untuk memodelkan hubungan non-linear yang kompleks dari fitur-fitur yang telah diekstraksi.
2. Untuk mendapatkan performa maksimal, penelitian di masa depan dapat menerapkan teknik optimisasi hiperparameter otomatis seperti Bayesian Optimization atau Genetic Algorithms untuk menemukan konfigurasi KAN (misalnya, grid_size, spline_order, struktur layer) yang paling optimal untuk setiap dataset.
3. Menguji ketahanan (robustness) model KAN terhadap data dengan tingkat noise yang tinggi akan memberikan wawasan tentang keandalannya dalam skenario dunia nyata yang seringkali tidak ideal.
4. Salah satu keunggulan utama KAN adalah interpretasinya. Penelitian lanjutan dapat fokus pada analisis mendalam terhadap bentuk fungsi spline yang dipelajari oleh KAN. Analisis ini berpotensi memberikan wawasan tentang hubungan kausal atau dinamis antar variabel atau antar langkah waktu (time lags) yang mungkin tidak terungkap oleh model lain.

DAFTAR PUSTAKA

- [1] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.
- [2] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- [3] De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3), 443-473.
- [4] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Time series forecasting with deep learning: a survey. *ACM Computing Surveys (CSUR)*, 54(7), 1-36.
- [5] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.
- [6] Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR)*.
- [7] Kitaev, N., Kaiser, L., & Levskaya, A. (2020). Reformer: The Efficient Transformer. *International Conference on Learning Representations (ICLR)*.
- [8] Kolmogorov, A. N. (1957). On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Doklady Akademii Nauk SSSR*, 114(5), 953-956.
- [9] Lai, G., Chang, W. C., Yang, Y., & Liu, H. (2018). Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- [10] Lim, B., & Zohren, S. (2021). Deep learning for time series forecasting: A survey. *Applied Sciences*, 11(11), 5122.
- [11] Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., & Tegmark, M. (2024). KAN: Kolmogorov-Arnold Networks. *arXiv preprint arXiv:2404.19756*.
- [12] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. *International Conference on Learning Representations (ICLR)*.
- [13] Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. *arXiv preprint arXiv:1609.03499*.

- [14] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- [15] Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *International Journal of Forecasting*, 36(3), 1181-1191.
- [16] Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90, 106181.
- [17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30.
- [18] Wen, Q., Torkkola, K., Narayanaswamy, B., & Madeka, D. (2017). A review on deep learning for time series forecasting. *arXiv preprint arXiv:1705.09412*.
- [19] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *Advances in Neural Information Processing Systems (NeurIPS)*, 34.
- [20] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 35(12), 11106-11115.