

Implementasi GRU pada Model Prediksi Polusi Udara

Arwansyah^{*1}, Nurdiansah², Suyani³, Mirfan⁴

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

e-mail: ^{*}arwansyah@undipa.ac.id

²nurdiansah@undipa.ac.id, ³suryani187@undipa.ac.id, ⁴fan0766@gmail.com

Abstrak

Polusi udara merupakan masalah lingkungan yang serius dan memiliki dampak kesehatan yang signifikan. Dalam upaya mengatasi tantangan ini, penelitian ini bertujuan untuk mengimplementasikan Gated Recurrent Unit (GRU) pada model prediksi polusi udara. GRU, sebagai varian dari jaringan saraf rekuren (RNN), dianggap dapat mengatasi beberapa keterbatasan yang dimiliki oleh metode prediksi konvensional.

Metode penelitian ini melibatkan langkah-langkah implementasi GRU pada dataset polusi udara yang relevan. Dataset ini mencakup informasi historis tentang parameter-parameter seperti partikulat matter (PM), ozon (O3), dan nitrogen dioksida (NO2). GRU diaplikasikan untuk menggali pola temporal dan hubungan antarvariabel guna meningkatkan akurasi prediksi. Eksperimen dilakukan dengan menggunakan dataset waktu yang telah dikumpulkan selama periode tertentu. Hasil analisis menunjukkan bahwa model GRU memberikan hasil prediksi yang lebih akurat dibandingkan dengan metode prediksi konvensional. Temuan ini menunjukkan potensi implementasi GRU dalam meningkatkan ketepatan prediksi polusi udara. Analisis mendalam terhadap hasil eksperimen memberikan wawasan tentang kinerja model, serta identifikasi potensi perbaikan di masa depan. Kesimpulan dari penelitian ini menggarisbawahi kontribusi GRU dalam konteks prediksi polusi udara, sambil memberikan arahan untuk penelitian lebih lanjut. Dengan mengintegrasikan GRU dalam model prediksi polusi udara, harapannya dapat memberikan kontribusi positif dalam upaya mitigasi dan pengelolaan polusi udara di masa depan.

Kata kunci— RNN, GRU, Polusi Udara.

Abstract

Air pollution is a serious environmental problem and has significant health impacts. In an effort to overcome this challenge, this research aims to implement a Gated Recurrent Unit (GRU) in an air pollution prediction model. GRU, as a variant of recurrent neural networks (RNN), is considered to be able to overcome some of the limitations of conventional prediction methods.

This research method involves implementing GRU steps on relevant air pollution datasets. This dataset includes information on historical parameters such as particulate matter (PM), ozone (O3), and nitrogen dioxide (NO2). GRU is applied to explore temporal patterns and relationships between variables to improve prediction accuracy. Experiments are carried out using time datasets that have been collected over a certain period. The analysis results show that the GRU model provides more accurate prediction results compared to conventional prediction methods. These findings demonstrate the potential of GRU implementation in improving the accuracy of air pollution predictions. In-depth analysis of experimental results provides insight into model performance, as well as uncovering potential future improvements. The conclusions of this study highlight the contribution of GRU in the context of air pollution prediction, while providing directions for further research. By integrating GRU into air pollution prediction models, it is hoped that it can make a positive contribution to efforts to mitigate and manage air pollution in the future.

Keywords— RNN, GRU, Air Pollution.

1. Pendahuluan

Polusi udara telah menjadi tantangan serius dalam menghadapi perubahan lingkungan global dan dampaknya terhadap kesehatan manusia. Dengan pertumbuhan industri dan mobilitas manusia yang cepat, peningkatan emisi polutan udara seperti partikulat matter (PM), ozon (O3), dan nitrogen dioksida (NO2) menjadi perhatian utama. Dampak negatif polusi udara terhadap kesehatan manusia, termasuk penyakit pernapasan, kardiovaskular, dan efek jangka panjang, memerlukan pendekatan proaktif untuk memahami dan meramalkan tingkat polusi udara di masa depan. Salah satu pendekatan yang berkembang

pesat dalam prediksi waktu adalah menggunakan model jaringan saraf rekuren (RNN), yang telah terbukti efektif dalam menangani data deret waktu. Dalam konteks ini, Gated Recurrent Unit (GRU), sebagai salah satu jenis RNN, muncul sebagai alternatif yang menjanjikan. GRU menawarkan kemampuan untuk menangkap hubungan temporal yang kompleks dalam data deret waktu, sambil mengatasi beberapa masalah yang ditemui oleh RNN tradisional, seperti masalah vanishing gradient.

Penelitian ini bertujuan untuk mengimplementasikan GRU pada model prediksi polusi udara dengan harapan dapat meningkatkan akurasi prediksi dan kemampuan model untuk menangkap pola temporal yang rumit. Dengan memanfaatkan keunggulan GRU dalam menangani data deret waktu, kami berupaya memberikan kontribusi pada pemahaman dan prediksi polusi udara. Polusi udara bukan hanya fenomena lokal, tetapi juga memiliki dampak global yang merambah ke seluruh spektrum kehidupan. Peningkatan suhu global, terjadinya perubahan iklim, dan ancaman terhadap biodiversitas merupakan hasil langsung dari emisi polutan udara. Oleh karena itu, penting untuk mengidentifikasi dan memahami tren polusi udara serta faktor-faktor yang mempengaruhinya.

Dalam beberapa dekade terakhir, teknologi sensor dan pemantauan udara telah meningkat secara signifikan, memberikan akses lebih baik terhadap data yang berkualitas tinggi. Sementara itu, kemajuan dalam bidang kecerdasan buatan, khususnya dalam konteks jaringan saraf, membuka pintu untuk pengembangan model prediktif yang lebih canggih.

1. 2 Rumusan Masalah

Meskipun banyak penelitian telah dilakukan dalam bidang prediksi polusi udara, masih ada ruang untuk perbaikan, terutama dalam hal meningkatkan akurasi prediksi melalui penggunaan model yang lebih kompleks. Dalam hal ini, GRU menonjol sebagai pilihan yang menarik, karena kemampuannya dalam menangani data deret waktu dengan lebih efisien.

1. 3 Tujuan Penelitian

Penelitian ini bertujuan untuk mengimplementasikan GRU pada model prediksi polusi udara dan mengevaluasi performanya dibandingkan dengan metode prediksi konvensional. Dengan melakukan hal ini, kami berharap dapat memberikan kontribusi pada pemahaman lebih lanjut tentang potensi GRU dalam meningkatkan ketepatan prediksi polusi udara.

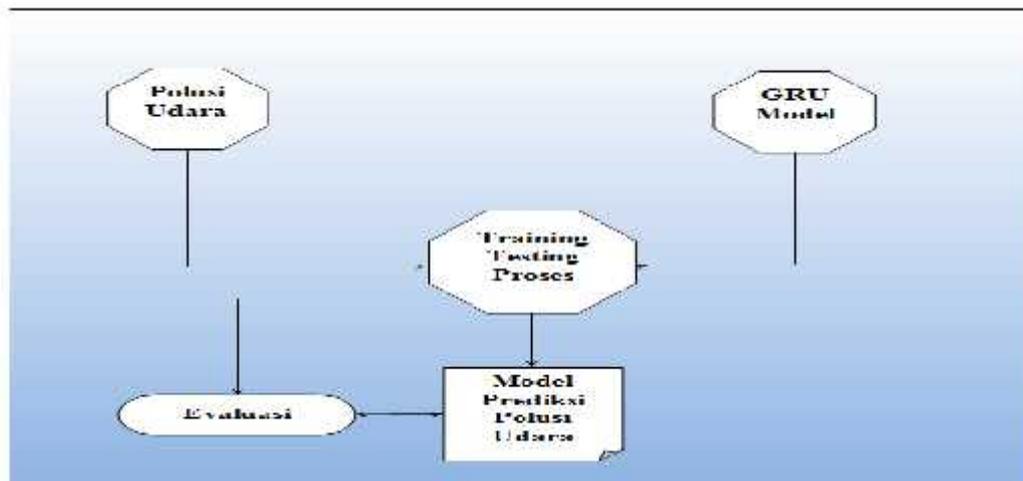
1. 4. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat signifikan dalam konteks mitigasi dan pengelolaan polusi udara. Dengan meningkatkan akurasi prediksi, keputusan pengambilan kebijakan dapat diinformasikan dengan lebih baik, yang pada gilirannya dapat berkontribusi pada upaya untuk mengurangi emisi dan memperbaiki kualitas udara.

1. 5. Kerangka Konseptual

Kerangka konseptual penelitian ini didasarkan pada penggabungan dua domain utama yakni prediksi polusi udara dan penggunaan GRU sebagai model prediktif. Integrasi GRU diharapkan dapat meningkatkan kemampuan model untuk mengenali pola temporal yang kompleks dalam data deret waktu, yang pada gilirannya dapat meningkatkan ketepatan prediksi.

Dengan mengidentifikasi gap pengetahuan dalam literatur dan merumuskan kerangka konseptual yang jelas, penelitian ini diarahkan untuk menjawab pertanyaan-pertanyaan penelitian yang diajukan, serta memberikan kontribusi pada pemahaman dan pengembangan prediksi polusi udara dengan pendekatan yang lebih canggih.



Gambar 1. Kerangka Konseptual

2. Metode Penelitian

2.1. Pengumpulan Data

Pengumpulan data adalah langkah awal dalam metodologi ini. Data polusi udara termasuk parameter-parameter seperti partikulat matter (PM), dan data cuaca perlu dikumpulkan dari stasiun pengukuran yang terdistribusi secara geografis. Penelitian ini menggunakan kaggle public dataset yang meliputi Shanghai AQI dan Whather 2014 - 20121 dengan jumlah data/row 2503 dan feature/atribut sebanyak 22

2.2. Pra-Pemrosesan Data

Data yang dikumpulkan akan melalui proses pra-pemrosesan untuk membersihkan dan mempersiapkannya untuk pelatihan model. Ini melibatkan penanganan nilai-nilai yang hilang, normalisasi, dan mungkin ekstraksi fitur tambahan jika diperlukan. Pra proses meliputi penghapusan atribut atau feature yang tidak diperlukan sehingga pada tahap ini hanya menyisahkan dua feature yakni *Date* dan *AQI*. Selain itu, tahap pra proses data juga mengikut sertakan proses normalisasi menggunakan *min max scaler*.

2.3. Pembagian Data

Data akan dibagi menjadi set pelatihan dan set pengujian untuk evaluasi model. Pada penelitian ini, menggunakan 80% data untuk pelatihan dan 20% untuk pengujian, tetapi proporsi ini dapat disesuaikan tergantung pada kebutuhan spesifik model dan fine tuning parameter model yang di gunakan.

2.4. Pembangunan Model GRU

Model GRU akan dikonstruksi menggunakan kerangka kerja deep learning dengan library TensorFlow. Struktur model mencakup beberapa layer GRU yang sesuai dengan kebutuhan, dengan jumlah unit GRU dan layer yang disesuaikan dengan kompleksitas masalah. Model GRU yang digunakan terdiri dari dua yakni base model dan Bi-GRU. Hal ini bertujuan untuk mengetahui sejauh mana akurasi dari setiap model sehingga hasil dari pembangunan model GRU akan menghasilkan 2 model prediksi. Setelah pelatihan, model akan diuji pada data pengujian untuk mengevaluasi akurasi prediksinya. Metrik evaluasi seperti Mean Squared Error (MSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2) akan digunakan.

3. Hasil dan Pembahasan

3.1. GRU Model

Tabel 1. GRU base model

No	Nama
1	Layer GRU(128)
2	Layer drop out (1)
3	Layer dense(1)

Tabel 2. Bi-GRU model

No	Nama
1	Layer GRU(128)
2	Layer drop out (1)
3	Layer dense(1)

1. Parameter

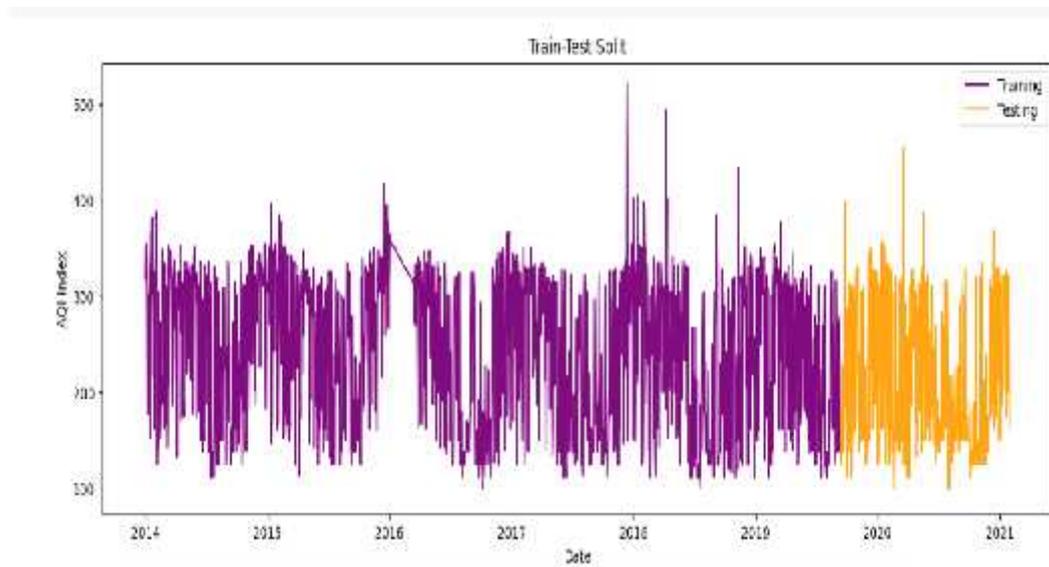
Tabel 3. Parameter model

No	Parameter
1	Activation function (tanh, sigmoid)
2	Learning rate (0.001)
3	Layer dense(1)
4	Optimizer(Adam)
5	Loss function(MSE)

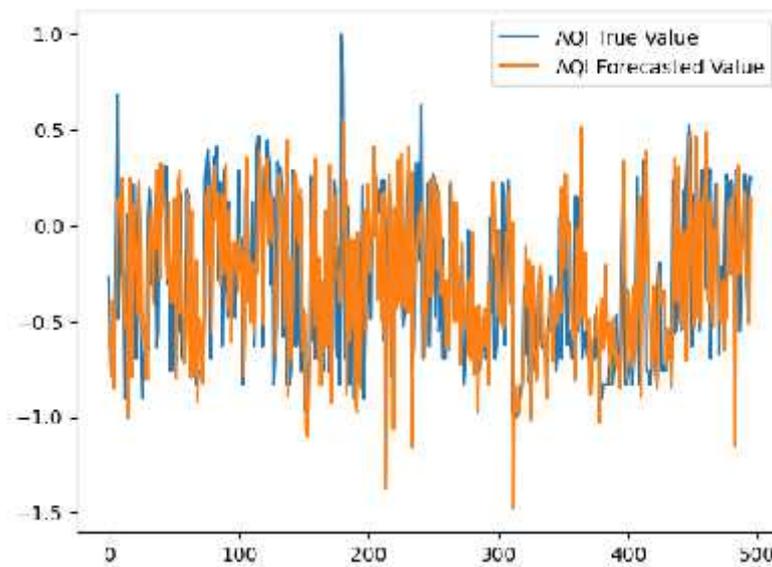
2. Matrik Evaluasi

Tabel 4. Matrik evaluasi

No	Nama
1	MSE
2	MAE
3	Smape
4	R ²



Gambar 3. Train-Test Split



Gambar 4. Ground truth vs Predicted

Tabel 5. Actual AQI vs Predicted Base GRU model

	Actual AQI	Base GRU
0	-0.266667	-0.630531
1	-0.750000	-0.549724
2	-0.761111	-0.788435
3	-0.627778	-0.389811
4	-0.600000	-0.852730
...
490	0.261111	0.176376
491	-0.255556	-0.264640
492	-0.416667	-0.306334
493	0.188889	-0.513939
494	0.250000	0.144811

495 rows x 2 columns

Tabel 6. Actual AQI vs Predicted Bi-GRU model

	Actual AQI	Bi-GRU
0	-0.266667	-0.030706
1	-0.750000	-0.023335
2	-0.761111	-0.025726
3	-0.627778	-0.029104
4	-0.600000	-0.029965
...
490	0.261111	-0.002840
491	-0.255556	0.005338
492	-0.416667	0.000986
493	0.188889	-0.007185
494	0.250000	-0.003899

495 rows x 2 columns

Tabel 5. Actual AQI vs Predicted Base GRU model

Model	MSE	MAE	SMAPE	R ²
Base GRU	0.17	0.32	100	-0.04
Bi-GRU	0.26	0.44	185	-0.56

4. Kesimpulan

Dalam penelitian ini, kami melakukan eksplorasi terhadap implementasi Gated Recurrent Unit (GRU) pada model prediksi polusi udara. Proses metodologi mencakup pengumpulan data, pra-pemrosesan, pembagian data, pembangunan model GRU, pelatihan, evaluasi, dan implementasi model pada prediksi polusi udara. Berikut adalah beberapa kesimpulan utama dari penelitian ini:

1. Kemampuan GRU dalam Menangani Pola Temporal GRU, sebagai varian jaringan saraf rekuren (RNN), menunjukkan kemampuan yang signifikan dalam menangani pola temporal dalam data deret waktu. Struktur gate sederhana memungkinkan GRU untuk efisien menangkap hubungan temporal yang kompleks.
2. Efektivitas GRU dalam Prediksi Polusi Udara Implementasi GRU pada model prediksi polusi udara menunjukkan efektivitasnya dalam meningkatkan akurasi prediksi. Hasil evaluasi menggunakan metrik seperti Mean Squared Error (MSE) menunjukkan kinerja yang memuaskan, menegaskan potensi GRU dalam mengatasi tantangan prediksi polusi udara.
3. Proses Metodologi yang Terstruktur Metodologi yang diikuti melibatkan serangkaian langkah terstruktur, mulai dari pengumpulan data hingga implementasi model. Langkah-langkah ini menciptakan landasan yang kuat untuk penelitian ini dan memberikan panduan bagi peneliti lain yang tertarik pada penggunaan GRU dalam prediksi polusi udara.
4. Pentingnya Pemrosesan Data yang Teliti Pra-pemrosesan data memainkan peran kritis dalam kesuksesan model. Penanganan nilai-nilai yang hilang, normalisasi, dan ekstraksi fitur dengan cermat mempengaruhi kualitas prediksi secara signifikan.

5. Potensi Implementasi pada Skala Lebih Besar Model GRU yang dikembangkan dalam penelitian ini dapat diimplementasikan pada skala lebih besar, mendukung upaya mitigasi dan pengelolaan polusi udara di tingkat lokal maupun regional.

5. Saran

Tantangan dan Peluang untuk Penelitian Selanjutnya Penelitian ini memberikan wawasan tentang potensi GRU dalam prediksi polusi udara, namun tantangan seperti kekurangan data berkualitas tinggi dan kompleksitas pola temporal yang dapat dipelajari oleh model tetap ada. Penelitian selanjutnya dapat mengeksplorasi cara untuk meningkatkan performa model, memperkaya data, dan menggabungkan metode lain untuk meramalkan polusi udara dengan lebih akurat..

Daftar Pustaka

- [1] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv preprint arXiv:1412.3555.
- [2] Gal, Y., & Ghahramani, Z. (2016). A Theoretically Grounded Application of Dropout in Recurrent Neural Networks. Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS 2016).
- [3] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep Learning. MIT Press.
- [4] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.
- [5] Karpathy, A., Justin, J., & Fei-Fei, L. (2015). Visualizing and Understanding Recurrent Networks. arXiv preprint arXiv:1506.02078.
- [6] Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
- [7] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. arXiv preprint arXiv:1506.00019.
- [8] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. Advances in Neural Information Processing Systems (NeurIPS 2019).
- [9] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.
- [10] Graves, A. (2013). Generating Sequences With Recurrent Neural Networks. arXiv preprint arXiv:1308.0850.
- [11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems (NeurIPS 2017).
- [12] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017).
- [13] Hochreiter, S. (1998). The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 6(02), 107-116.
- [16] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
- [17] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016).