

Estimasi Utilisasi Prosesor pada Jaringan Interkoneksi Optik menggunakan Regresi Gaussian

HILAL HUDAN NUHA¹, AULIA ARIF WARDANA²

¹HUMIC Engineering, Telkom University, Indonesia

²Information and Communication Technology, Wrocław University of Science and Technology, Poland

Email: hilalnuha@telkomuniversity.ac.id

Received 24 Februari 2022 | *Revised* 27 Maret 2022 | *Accepted* 22 April 2022

ABSTRAK

Teknologi prosesor paralel melibatkan lebih dari satu node prosesor dalam interkoneksi optik. Unjuk kerja dari desain dari jaringan optik tersebut harus diestimasi sebelum diimplementasikan. Atribut yang bisa diambil diambil dari desain tersebut yaitu nomor node dan thread, distribusi spasial dan temporal, dan rasio T/R. Kajian ini mengevaluasi teknik estimasi utilitas prosesor menggunakan regresi Gaussian yang dibandingkan dengan Support Vector Machine untuk regresi dan regresi linear. Hasil percobaan penunjukkan bahwa regresi Gaussian menghasilkan akurasi estimasi paling tinggi dengan nilai koefisien determinasi sebesar 98.75%.

Kata kunci: *Utilisasi prosesor, interkoneksi optik, regresi Gaussian*

ABSTRACT

Parallel processor technology involves more than one processor node in an optical interconnection. The performance of the design of the optical network must be estimated before further deployment. The attributes that can be obtained from the design are the number of nodes and threads, the spatial and temporal distribution, and the T/R ratio. This study evaluates the processor utility estimation technique using Gaussian regression compared to the Support Vector Machine for linear regression and regression. The experimental results show that Gaussian regression produces the highest estimation accuracy with a coefficient of determination of 98.75%.

Keywords: *Processor Utilization, Optical Interconnection, Gaussian Regression*

1. PENDAHULUAN

Teknologi prosesor paralel membuat berbagai persoalan yang melibatkan data dalam jumlah besar dan independen bisa dilakukan secara cepat. Kemunculan Intel 80286 dan Sun SPARC pada 1970an membuat tren teknologi komputasi *parallel* semakin meningkat (Healey dkk, 2020). Namun desain dari prosesor *parallel* yang rumit membuat estimasi unjuk kerja dari prosesor sulit untuk dilakukan menggunakan pendekatan analitik (Akay, dkk, 2015). Oleh karena itu pendekatan empiris yang dimulai dari data yang diambil banyak dilakukan (Zayid & Akay, 2013)(Akay & Abasikeleş 2010). Pendekatan empiris tersebut membutuhkan sampel yang harus dianalisis menggunakan metode statistik atau metode berbasis pembelajaran mesin yang membantu memodelkan korelasi antara masukan dan luaran sehingga bisa melakukan estimasi yang akurat.

Salah satu pendekatan empiris yang digunakan untuk melakukan estimasi performansi prosesor adalah regresi Gaussian. Metode ini sangat efektif karena tidak membutuhkan pemilihan parameter yang rumit. Salah satu pembahasan tentang regresi Gaussian untuk pemodelan bisa dilihat di makalah (Schulz, dkk, 2018). Tantangan utama dari penggunaan regresi Gaussian pada ilmu material bisa dilihat pada makalah (Deringer, dkk, 2021) (Iannace & Ciaburro, 2021)(Ballabio, dkk, 2019). Oleh karena itu, teknik regresi Gaussian telah menarik perhatian para peneliti. Penelitian terkait penggunaan *Hilbert space* untuk regresi Gaussian juga bisa dilihat di artikel (Solin & Särkkä, 2020). Aplikasi dari metode regresi Gaussian untuk deteksi status baterai juga dibahas lebih mendalam pada karya ilmiah (Richardson, dkk, 2017)(Li, dkk, 2020)(Li, dkk, 2020). Makalah (Ghasemim, dkk, 2021) juga menggunakan teknik regresi Gaussian untuk penggunaan prediksi musim kering. Metode regresi Gaussian juga dipakai untuk penggunaan estimasi kandungan sulfat di karya ilmiah (Zhao, dkk, 2019). Studi tentang regresi Gaussian untuk estimasi parameter juga bisa dilihat di makalah (Mardalena, dkk, 2020). Kajian implementasi dari regresi Gaussian untuk estimasi usia bisa dibahas pada makalah (Sawant & Bhurchandi, 2019). Selain itu, makalah (Kong, dkk, 2018) menyajikan regresi Gaussian untuk asesmen *tool wear*. Regresi Gaussian juga digunakan untuk estimasi jumlah hasil produksi panen (Alebele, dkk, 2021). Estimasi kondisi turbin angin menggunakan regresi Gaussian dikaji pada makalah (Avendaño-Valencia, dkk, 2021). Karya-karya ilmiah tersebut menunjukkan bahwa metode regresi Gaussian cocok digunakan untuk berbagai permasalahan (Li, dkk, 2020).

Estimasi performansi dari prosesor paralel menggunakan teknik pembelajaran mesin seperti pada (Zayid & Akay, 2013a; 2013b). Teknik *fuzzy neural network* (de Campos Souza, dkk, 2021) juga dipakai untuk estimasi unjuk kerja dari interkoneksi optik. Namun teknik regresi Gaussian belum dijumpai untuk keperluan tersebut. Kontribusi utama dari kajian ini adalah analisis penggunaan regresi Gaussian untuk estimasi utilisasi prosesor pada interkoneksi optic. Struktur kajian ini terdiri atas lima bagian. Bagian I memberikan pengantar tentang permasalahan estimasi utilisasi prosesor. Metodologi yang dipakai dan regresi Gaussian dibahas pada Bagian II. Perancangan dan percobaan regresi Gaussian disampaikan pada Bagian III. Kesimpulan diberikan pada Bagian IV.

2. METODE PENELITIAN

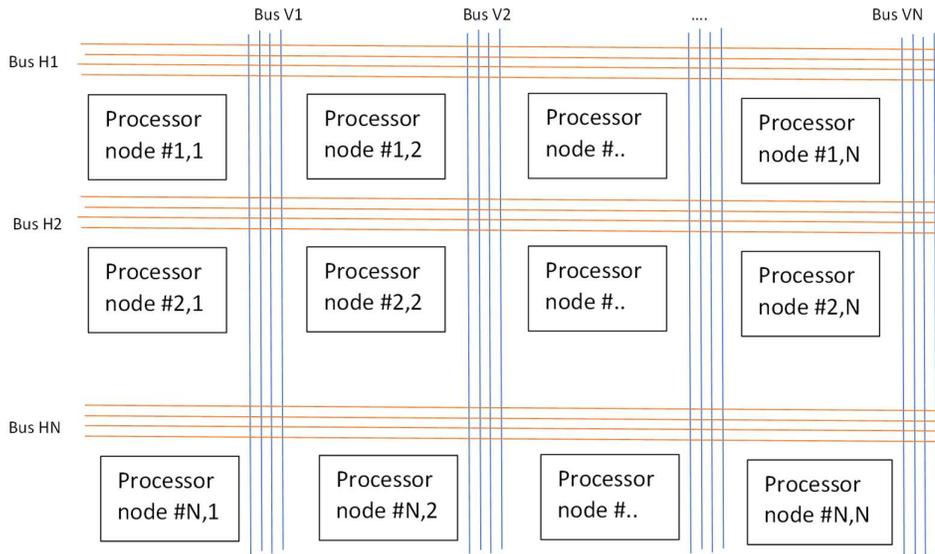
Bagian ini membahas arsitektur prosesor, atribut dari data, regresi Gaussian, dan pengukuran akurasi.

2.1 Arsitektur Prosesor

Makalah ini menggunakan arsitektur *2-Dimensional Simultaneous Optical Multiprocessor Exchange* (2D SOME)-*Bus*. Prosesor ini terdiri atas $N \times N$ buah node prosesor dimana masing-

masing *node* prosesor terhubung ke satu buah bus vertikal dan satu buah bus *horizontal* seperti yang ditunjukkan pada Gambar 1.

Setiap *node* dua pasang pemancar dan penerima pesan yang memungkinkan pengiriman pesan pada suatu bus ke semua *node*. Setiap pesan memuat header yang memuat informasi node tujuan yang disebarikan pada bus yang dituju. *Node* sumber bisa memilih untuk menyebarkan pesan pertama pada *node* perantara melalui bus vertikal yang selanjutnya disebarikan ulang melalui bus *horizontal*.



Gambar 1. Arsitektur Prosesor Optik

2.2 Atribut dan Data

Makalah ini menggunakan data yang dihasilkan oleh Akay dkk (**Akay, dkk, 2015**) yang memodelkan arsitektur pada Gambar 1 menggunakan OPNET yang menghasilkan 640 sampel. Data yang digunakan mempunyai atribut masukan berupa *Node Number*, *Thread Number*, *Spatial Distribution*, *Temporal Distribution*, dan T/R yang merupakan rasio durasi transfer dengan durasi *runtime*. Sedangkan luaran dari data tersebut yaitu utilisasi prosesor (*Processor Utilization*) yang menggambarkan tingkat penggunaan prosesor.

Distribusi nilai dari atribut masukan dan luaran tersebut ditunjukkan oleh Gambar 2. Atribut nomor *node* dan *thread* dan distribusi spasial dan temporal mempunyai sebaran yang merata. Namun atribut T/R mempunyai sebaran yang lebih beragam dengan kekosongan nilai interval T/R 0.3-0.4 dan 0.6-0.7. Target luaran (nilai utilisasi prosesor) mempunyai sebaran beragam yaitu dengan utilisasi pada interval 0.5-0.6 mempunyai frekuensi kemunculan yang paling tinggi. Semua parameter tersebut tidak mempunyai satuan karena mencerminkan nomor identitas node maupun hasil bagi dari dua nilai dengan satuan yang sama.

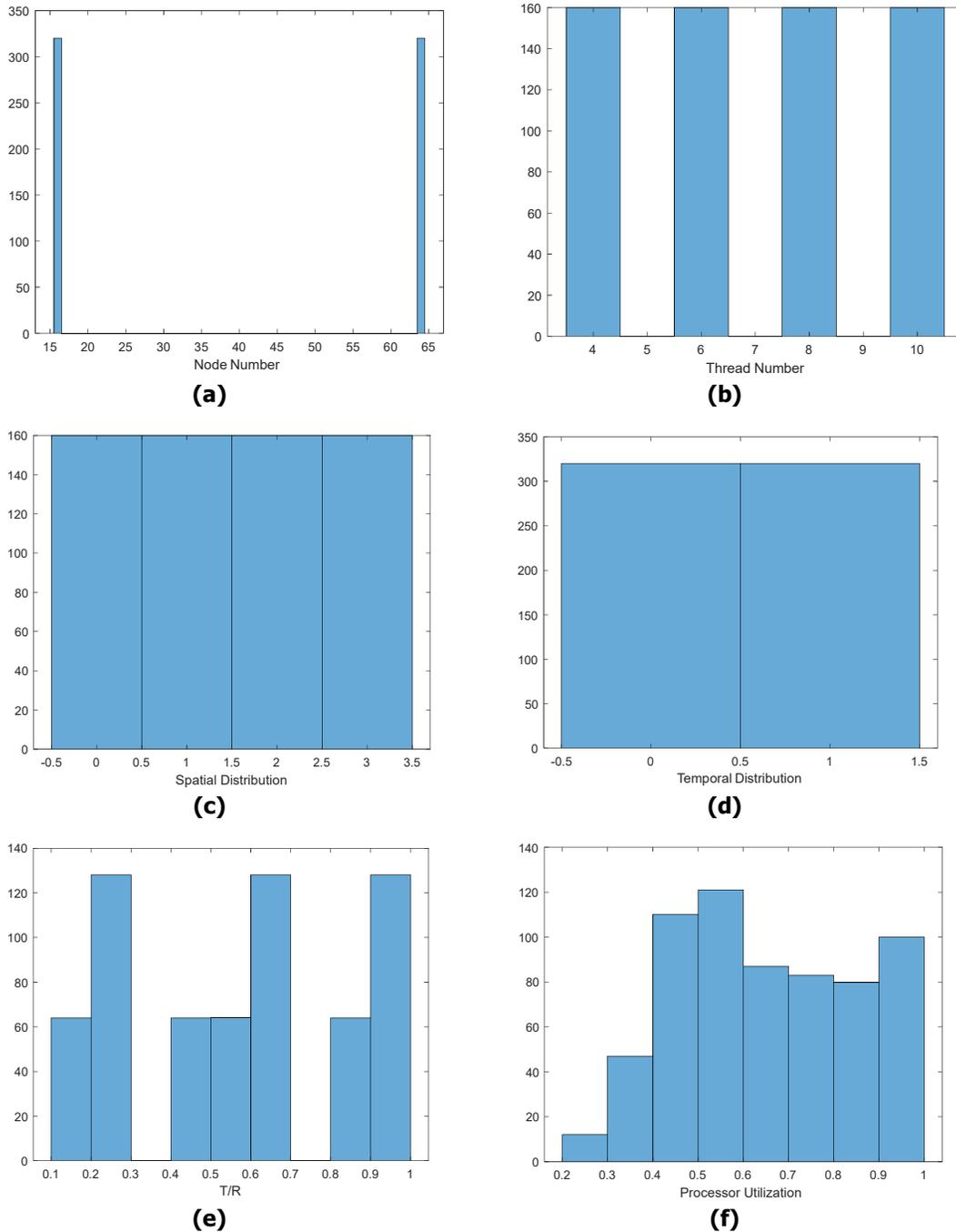
2.3 Regresi Gaussian

Regresi Gaussian (RG) memodelkan atribut masukan x dan luaran o sesuai Persamaan (1):

$$o = x\beta + \epsilon, \tag{1}$$

dimana β adalah koefisien dari data yang akan diestimasi dan ϵ adalah variabel acak yang diambil dari distribusi normal $N(0, \sigma^2)$ (**"Gaussian Process Regression Models" n.d.**). Pada makalah ini, varian dari variabel acak tersebut menggunakan nilai $\sigma = 2$. Pemilihan nilai σ

mempengaruhi akurasi dari estimasi. Namun, dalam pemilihan nilai σ tersebut, penentuan nilai yang optimal bisa dilakukan secara manual dalam beberapa kali ujicoba (<5).



Gambar 2. Histogram sebaran nilai pada data berdasar atribut dan luaran

2.4 Pengukuran Akurasi

Makalah ini menggunakan tiga buah metrik pengukuran estimasi yaitu *Root Mean Squared Error (RMSE)*, *Mean Absolute Percentage Error (MAPE)*, dan koefisien determinasi (R^2). Ketiga metrik tersebut dihitung dengan persamaan berikut:

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (t_n - o_n)^2}{N}} \quad (2)$$

$$MAPE = \frac{1}{N} \sum_{n=1}^N \left| \frac{t_n - o_n}{t_n} \right| \times 100\% \quad (3)$$

$$R^2 = \left(1 - \frac{\sum_{n=1}^N (t_n - \bar{t})^2}{\sum_{n=1}^N (t_n - o_n)^2} \right) \times 100\% \quad (4)$$

Dimana o adalah luaran hasil estimasi dan t adalah nilai yang sebenarnya. Ketiga metrik tersebut biasa digunakan untuk mengukur akurasi hasil prediksi sebuah model. RMSE dan MAPE mencerminkan selisih antara hasil estimasi dan nilai yang sebenarnya. Sedangkan R^2 mencerminkan seberapa akurat model yang dibuat untuk menghasilkan luaran yang sesuai dengan nilai aktual.

3. HASIL DAN PEMBAHASAN

Percobaan dilakukan menjadi tiga bagian utama yaitu pengukuran pengaruh proporsi data terhadap akurasi, perbandingan dengan *Support Vector Machine* (SVM) dan regresi linear (RL), dan pengaruh atribut terhadap akurasi. Eksperimen dilakukan dengan Matlab 2021a menggunakan program yang bisa diakses pada tautan berikut <https://github.com/hilalnuha/ProcessorGaussEstimator>.

3.1 Pengaruh jumlah data latih terhadap tingkat akurasi

Pada bagian ini, percobaan dilakukan untuk melihat pengaruh jumlah data training μ terhadap akurasi. Tabel 1. menunjukkan pengaruh jumlah data latih terhadap nilai RMSE. Tabel tersebut menunjukkan bahwa secara rata-rata dan nilai minimum RMSE mengalami penurunan ketika jumlah data latih diperbanyak. Nilai maksimum RMSE cenderung tetap sekalipun jumlah data latih ditambah.

Tabel 1. Pengaruh jumlah data training terhadap nilai RMSE

RMSE	$\mu=50\%$	$\mu=75\%$	$\mu=80\%$	$\mu=90\%$
Min	0.027	0.027	0.025	0.023
Max	0.032	0.033	0.033	0.033
Mean	0.030	0.031	0.028	0.027

Pola yang sama juga ditunjukkan oleh Tabel 2 dimana nilai MAPE semakin baik ketika semakin banyak data latih yang dipakai. Sama seperti RMSE, nilai MAPE minimum dan rata-rata turun ketika jumlah data latih ditingkatkan. Nilai MAPE maksimum juga cenderung turun kecuali pada porsi data latih tertinggi yang sedikit lebih tinggi dari pada porsi data latih di bawahnya.

Tabel 2. Pengaruh jumlah data training terhadap nilai MAPE

MAPE (%)	$\mu=50\%$	$\mu=75\%$	$\mu=80\%$	$\mu=90\%$
Min	4.43	3.34	2.96	2.61
Max	4.76	3.95	3.43	3.52
Mean	4.60	3.62	3.18	3.06

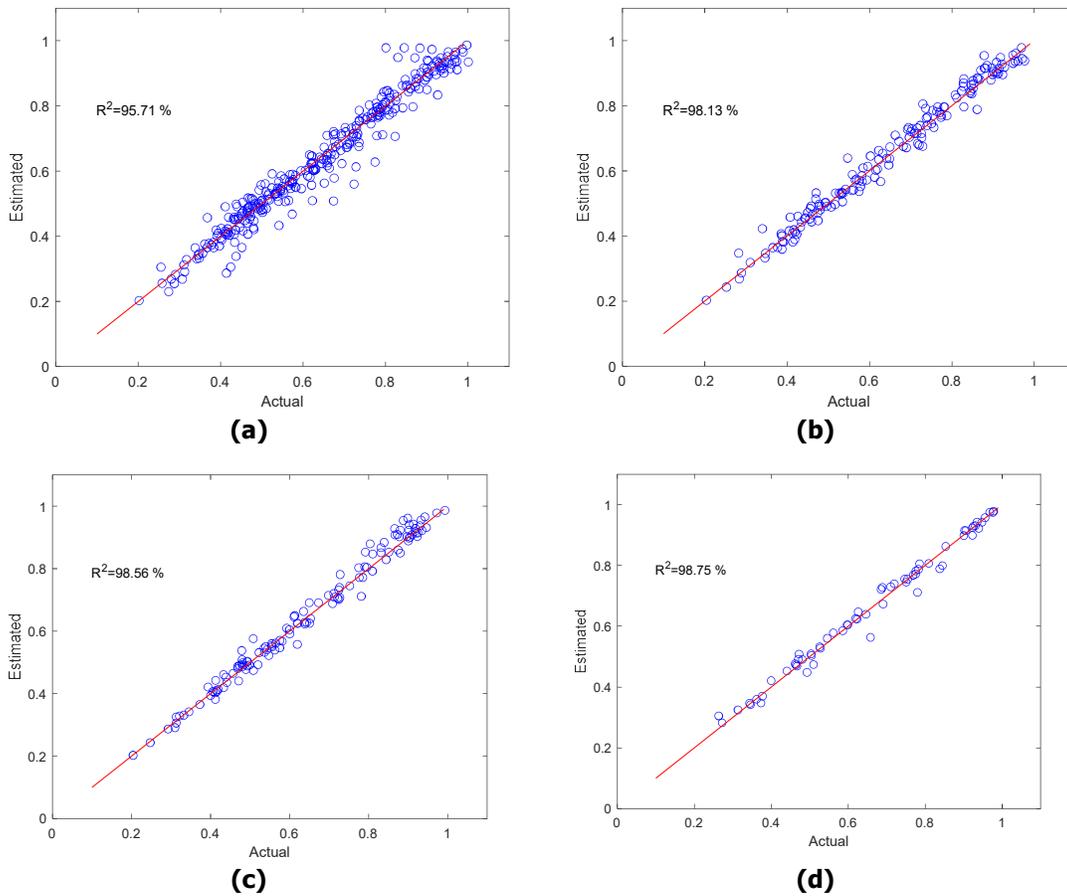
Koefisien determinasi (Tabel 3) yang mencerminkan tingkat korelasi prediksi juga menunjukkan bahwa jumlah data latih meningkatkan unjuk kerja dari metode. Nilai koefisien determinasi maksimum cenderung naik secara konsisten walaupun dari sisi rata-rata dan nilai

minimum koefisien ini tidak mengalami kenaikan yang konsisten. Secara umum bisa disimpulkan bahwa penggunaan data latih meningkatkan akurasi dari metode estimasi.

Tabel 3. Pengaruh jumlah data training terhadap nilai R²

R ² (%)	$\mu=50\%$	$\mu=75\%$	$\mu=80\%$	$\mu=90\%$
Min	94.75	97.12	97.06	96.79
Max	95.71	98.13	98.56	98.75
Mean	95.23	97.54	98.04	98.01

Tingkat korelasi dari hasil prediksi dan nilai acuan bisa dilihat pada Gambar 3 berupa *scatter plot*. Gambar tersebut diambil dari hasil pengujian yang menghasilkan nilai koefisien determinasi tertinggi pada setiap kasus. Jumlah data pada setiap panel pada gambar tersebut berkurang karena porsi data uji yang menurun. Semakin tinggi nilai koefisien maka titik plot akan semakin mendekati garis diagonal. Begitu juga sebaliknya, semakin rendah nilai koefisien determinasi akan menghasilkan titik plot yang menyebar. Titik plot yang berada di bawah garis diagonal mencerminkan bahwa hasil estimasi lebih rendah dari nilai aktual. Begitu juga sebaliknya, titik plot di atas garis diagonal merepresentasikan hasil estimasi yang lebih tinggi dari pada nilai aktual. Percobaan-percobaan tersebut menandakan bahwa akurasi dari RG untuk estimasi menghasilkan nilai yang bagus baik untuk RMSE, MAPE, maupun R².



Gambar 3. Nilai R² untuk (a) $\mu=50\%$, (b) $\mu=75\%$, (c) $\mu=80\%$, dan (d) $\mu=90\%$

3.2 Perbandingan Dengan Metode Lain

Pada bagian ini, RG akan dibandingkan dengan metode lain yaitu SVM untuk regresi dan RL menggunakan $\mu=90\%$. Tabel 4 menunjukkan bahwa RG secara signifikan menghasilkan RMSE=0.027 lebih kecil dibandingkan dengan SVM dan RL. Nilai error berupa RMSE dari metode RL menunjukkan bahwa nilai hubungan input dan output tidak bisa dimodelkan secara linear. Baik RL dan SVM menghasilkan nilai minimum RMSE yang sama. Dalam aspek nilai maksimum dan rata-rata RMSE, SVM mempunyai hasil yang lebih buruk dari pada RL.

Tabel 4. Nilai RMSE dari tiap metode

RMSE	RG	SVM	RL
Min	0.023	0.101	0.101
Max	0.033	0.154	0.153
Mean	0.027	0.128	0.126

Dalam hal MAPE, RG juga mendapatkan persen error yang jauh lebih kecil dari metode pembandingan seperti pada Tabel 5. Nilai MAPE minimum dari RV lebih tinggi dari RL walaupun nilai MAPE maksimum dan rata-rata SVM menghasilkan nilai yang lebih rendah dari pada RL.

Tabel 5. Nilai MAPE dari tiap metode

MAPE (%)	RG	SVM	RL
Min	2.61	13.84	13.55
Max	3.52	20.84	20.90
Mean	3.06	17.08	17.79

Nilai koefisien determinasi pada Tabel 6 juga menunjukkan bahwa hasil prediksi dari RG mendapatkan nilai terbaik. Dari aspek nilai minimum, SVM menghasilkan nilai yang lebih baik dari RL walaupun koefisien rata-rata dan maksimum SVM lebih rendah dari pada RL. Ketiga parameter tersebut secara jelas menunjukkan keunggulan dari RG untuk estimasi nilai utilisasi.

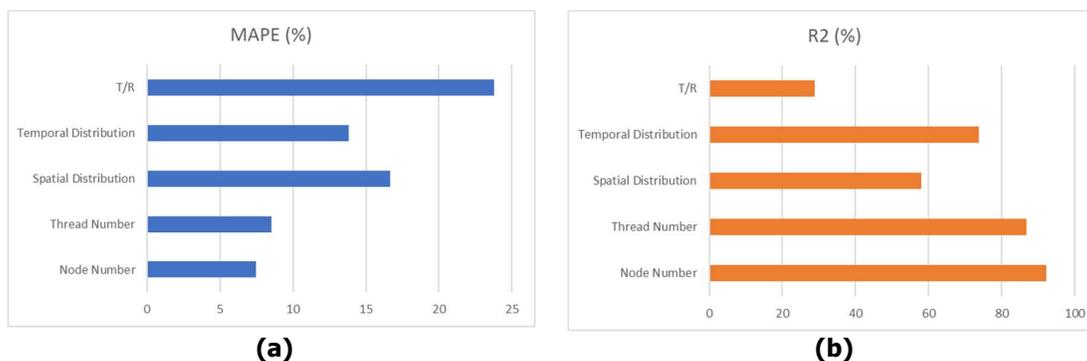
Tabel 6. Nilai R² dari tiap metode

R ² (%)	RG	SVM	RL
Min	96.79	46.69	48.29
Max	98.75	73.28	73.40
Mean	98.01	58.81	59.77

RG mempunyai akurasi yang tinggi karena fleksibilitas dari model untuk memodelkan data. Nilai σ membantu model dalam meningkatkan generalisasi dari sistem sehingga tidak terjadi *overfitting*. Oleh karena itu, nilai akurasi dari RG lebih tinggi dari SVM dan RL yang bisa dengan mudah terjebak pada *overfitting* pada data latih.

3.3 Pengaruh Atribut terhadap Akurasi

Pada bagian ini, salah satu atribut dalam estimasi dihilangkan sebagai masukan dari RG sehingga hanya terdiri atas empat atribut. Gambar 4 menunjukkan pengaruh atribut terhadap akurasi estimasi. Gambar 4(b) memperlihatkan bahwa nilai MAPE naik secara signifikan ketika atribut R/T dihilangkan diikuti oleh distribusi spasial, distribusi temporal, nomor *thread*, dan nomor *node*.



Gambar 4. Pengaruh tiap atribut pada (a) MAPE dan (b) R²

Nilai koefisien determinasi juga menunjukkan pola yang sama yaitu turunya koefisien determinasi secara drastis jika T/R tidak dimasukkan dalam estimasi. Kedua performansi tersebut menunjukkan bahwa T/R merupakan atribut paling berpengaruh dalam estimasi utilisasi jaringan. Secara umum atribut yang paling berpengaruh setelah T/R yaitu distribusi spasial, distribusi temporal, nomor *thread*, dan nomor *node*.

4. KESIMPULAN

Makalah ini mengkaji penggunaan regresi Gaussian untuk estimasi utilisasi prosesor pada jaringan interkoneksi optik. Atribut dari system seperti nomor node dan thread, distribusi spasial dan temporal, dan rasio T/R digunakan sebagai masukan dari regresi Gaussian dengan target luaran berupa utilisasi prosesor. Hasil percobaan menunjukkan bahwa regresi Gaussian menghasilkan estimasi paling akurat dibandingkan SVM dan regresi linear dalam aspek MAPE dan koefisien determinasi. Selain itu, percobaan yang dilakukan juga menunjukkan bahwa atribut T/R merupakan atribut paling berpengaruh dalam estimasi utilisasi prosesor. Hal tersebut disebabkan karena regresi Gaussian mempunyai parameter derau yang memungkinkan tercapainya akurasi dan kemampuan generalisasi yang tinggi.

DAFTAR RUJUKAN

- Akay, M. F., Aci, Ç. I., & Abut, F. A. T. İ. H. (2015). Predicting the performance measures of a 2-dimensional message passing multiprocessor architecture by using machine learning methods. *Neural Network World*, 25(3), 241.
- Akay, M. F., & Abasikeleş, I. (2010). Predicting the performance measures of an optical distributed shared memory multiprocessor by using support vector regression. *Expert Systems with Applications*, 37(9), 6293-6301.
- Alebele, Y., Wang, W., Yu, W., Zhang, X., Yao, X., Tian, Y., ... & Cheng, T. (2021). Estimation of Crop Yield From Combined Optical and SAR Imagery Using Gaussian Kernel Regression. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 10520-10534.

- Avendaño-Valencia, L. D., Abdallah, I., & Chatzi, E. (2021). Virtual fatigue diagnostics of wake-affected wind turbine via Gaussian Process Regression. *Renewable Energy*, *170*, 539-561.
- Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., ... & Panagos, P. (2019). Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma*, *355*, 113912.
- de Campos Souza, P. V., Soares, E. A., Guimarães, A. J., Araujo, V. S., Araujo, V. J. S., & Rezende, T. S. (2021). Autonomous data density pruning fuzzy neural network for optical interconnection network. *Evolving Systems*, *12*(4), 899-911.
- Deringer, V. L., Bartók, A. P., Bernstein, N., Wilkins, D. M., Ceriotti, M., & Csányi, G. (2021). Gaussian process regression for materials and molecules. *Chemical Reviews*, *121*(16), 10073-10141.
- Gaussian Process Regression Models*. n.d. Accessed February 16, 2022. <https://www.mathworks.com/help/stats/gaussian-process-regression-models.html>.
- Ghasemi, P., Karbasi, M., Nouri, A. Z., Tabrizi, M. S., & Azamathulla, H. M. (2021). Application of Gaussian process regression to forecast multi-step ahead SPEI drought index. *Alexandria Engineering Journal*, *60*(6), 5375-5392.
- Iannace, G., & Ciaburro, G. (2021). Modelling sound absorption properties for recycled polyethylene terephthalate-based material using Gaussian regression. *Building Acoustics*, *28*(2), 185-196.
- Kong, D., Chen, Y., & Li, N. (2018). Gaussian process regression for tool wear prediction. *Mechanical systems and signal processing*, *104*, 556-574.
- Li, X., Yuan, C., Li, X., & Wang, Z. (2020). State of health estimation for Li-Ion battery using incremental capacity analysis and Gaussian process regression. *Energy*, *190*, 116467.
- Li, X., Yuan, C., & Wang, Z. (2020). Multi-time-scale framework for prognostic health condition of lithium battery using modified Gaussian process regression and nonlinear regression. *Journal of Power Sources*, *467*, 228358.
- Mardalena, S., Purhadi, P., Purnomo, J. D. T., & Prastyo, D. D. (2020). Parameter estimation and hypothesis testing of multivariate Poisson inverse Gaussian regression. *Symmetry*, *12*(10), 1738.
- Healey, R., Dowers, S., Gittings, B., & Mineter, M. J. (Eds.). (2020). *Parallel processing algorithms for GIS*. CRC Press.
- Richardson, R. R., Osborne, M. A., & Howey, D. A. (2017). Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, *357*, 209-219.
- Sawant, M. M., & Bhurchandi, K. (2019). Hierarchical facial age estimation using Gaussian

process regression. *IEEE Access*, 7, 9142-9152.

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85, 1-16.

Solin, A., & Särkkä, S. (2020). Hilbert space methods for reduced-rank Gaussian process regression. *Statistics and Computing*, 30(2), 419-446.

Zayid, E. I. M., & Akay, M. F. (2013). Predicting the performance measures of a message-passing multiprocessor architecture using artificial neural networks. *Neural Computing and Applications*, 23(7), 2481-2491.

Zayid, E. I. M., & Akay, M. F. (2013b, August). Reliable attributes selection technique for predicting the performance measures of a dsm multiprocessor architecture. In 2013 *International Conference On Computing, Electrical And Electronic Engineering (ICCEEE)* (pp. 209-215). IEEE.

Zhao, J., Guo, H., Han, M., Tang, H., & Li, X. (2019). Gaussian process regression for prediction of sulfate content in lakes of China. *Journal of Engineering and Technological Sciences*, 51(2), 198-215.