

Analisis Prediksi Okupansi Jumlah Penumpang Kereta Api dengan Metode Support Vector Regression dan Gaussian Process Regression (Studi Kasus: Kereta Api Argo Parahyangan)

Ig. Prasetya Dwi Wibawa¹, Meta Kallista²

¹ Teknik Elektro, Universitas Telkom, Bandung, Indonesia.

Email: prasdwiwawa@telkomuniversity.ac.id

² Teknik Komputer, Universitas Telkom, Bandung, Indonesia.

Email: metakallista@telkomuniversity.ac.id

Manuscript submitted : October 2020

Accepted for publication : November 2020

Abstract: SVR (support vector regression) and GPR (gaussian process regression) are several methods in machine learning that are commonly used to accommodate regression problems. SVR and GPR have advantages over conventional regression methods. Both methods are non-parametric models, where learning models are built using mathematical functions. As a case study, we have predicted passenger occupancy of Argo Parahyangan Railway operated by PT Kereta Api Indonesia (Persero), which travel from Bandung to Gambir and vice versa. The research was conducted using occupancy data of passengers per day in year of 2019, especially economy class and executive class. Test scenarios are performed by comparing several metrics evaluations such as RMSE, R-squared, MAE, prediction speed, and computation time in the training and testing process using SVR and GPR methods.

2010 Mathematical Subject Classification: 62G08, 62J02, 68T05.

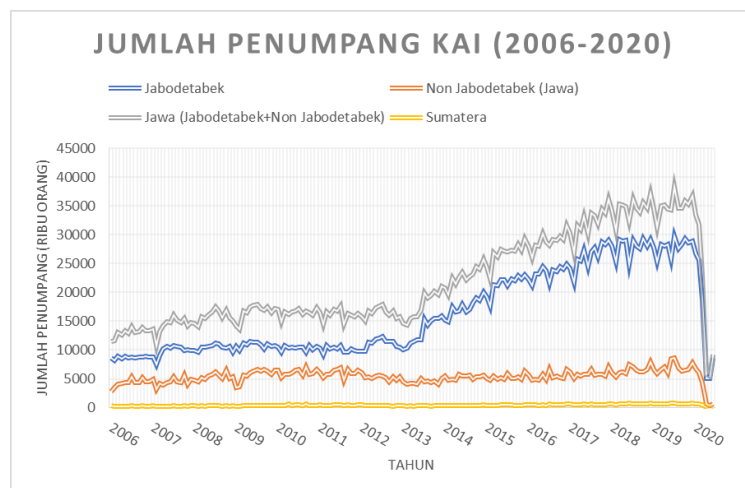
Keywords: SVR, GPR, prediction, railways, regression.

1. Pendahuluan

Dari data Biro Pusat Statistik (BPS) 2020, sejak awal tahun 2013 hingga akhir tahun 2019, peningkatan penumpang kereta api terus meningkat dari tahun ke tahun seperti terlihat pada Gambar 1. Peningkatan tersebut cukup signifikan sebelum terjadi pandemik Covid-19, yang mengharuskan masyarakat untuk melakukan pembatasan sosial berskala besar (PSBB). Peningkatan jumlah penumpang tersebut harus dipandang sebagai suatu analisis terhadap kebijakan PT Kereta Api Indonesia (KAI) dalam meningkatkan pelayanan terhadap konsumen. Di dalam makalah Jamaludin, 2019 [1], total biaya operasi untuk mengoperasikan satu armada kereta api lokal jarak dekat, dalam satu rangkaian perjalanan satu hari, mencapai kisaran 17.9 milyar rupiah. Di dalam hari libur atau hari raya, jumlah penumpang meningkat lebih

banyak. Hal ini tentunya menjadi studi yang menarik untuk dapat memprediksi level okupansi dari jumlah penumpang dalam rentang beberapa waktu ke depan sehingga dapat menjadi salah satu masukan informasi di dalam pengambilan kebijakan terkait aspek bisnis.

Salah satu metode di dalam pembelajaran mesin untuk melakukan prediksi, dalam kasus ini regresi, yaitu *Support Vector Regression (SVR)*. Salah satu kelebihan dari algoritme SVR adalah model pembelajarannya dapat mengakomodasi masalah fungsi non linier, dimana pendekatan regresi fungsi linier tidak dapat mengatasi permasalahan tersebut, sebagai contoh yaitu masalah regresi di dalam ruang fitur berdimensi tinggi. Adapun masalah optimasi diselesaikan secara numerik menggunakan pemrograman dual-konveks kuadratik. Beberapa studi literatur tentang prediksi level okupansi jumlah penumpang kereta api telah dilakukan antara lain G. Vandewiele, 2017 [2], melakukan prediksi okupansi kereta berdasarkan log antrian dan sumber data eksternal. Selain itu E. Jenelius, 2019 [3], melakukan prediksi keramaian kereta metro berbasis data berdasarkan data beban waktu-nyata. J. Guo, 2019 [4], melakukan prediksi arus penumpang abnormal jangka pendek menggunakan kombinasi SVR dan LSTM (*long short-term memory*). Pada tahun berikutnya, J. Guo, 2020 [5], mengembangkan algoritme pengenalan arus penumpang kereta api untuk stasiun terminal berdasarkan teori biaya dan kontrol frekuensi otomatis. B. Wang, 2020 [6], melakukan prediksi arus penumpang jangka pendek untuk stasiun kereta api perkotaan menggunakan jaringan pembelajaran berdasarkan algoritme input informasi arus penumpang yang optimal. Sementara Z. Jing, 2020 [7], melakukan studi eksplorasi terkait Model Prediksi Berbasis Jaringan Saraf untuk Arus Penumpang di Stasiun Penumpang Besar. Selain itu W. Lu, 2020 [8], melakukan penelitian tentang pemilihan sampel prediksi arus penumpang transit kereta api perkotaan berdasarkan algoritma SCBP. Selain masalah prediksi okupansi, beberapa masalah lain di dalam perkeretaapian antara lain masalah penjadwalan. J. Lei, 2019, memaparkan tentang pengaruh kedatangan kereta api dengan trafik moda transportasi lain untuk optimasi penjadwalan [9]. Sementara Y. Sun, 2019 [10], menggunakan metode hibrid untuk memprediksi umur relai kereta api berdasarkan dekomposisi multi-layer dan RBFNN (*radial basis function neural network*).



Gambar 1. Jumlah penumpang KAI periode Januari 2006 - Juni 2020.

Di dalam makalah ini akan coba dilakukan prediksi okupansi penumpang kereta api khususnya Kereta Api Argo Parahyangan dengan menggunakan model pembelajaran berbasis SVR dan juga GPR. SVR memiliki kekeluasaan di dalam menentukan seberapa banyak kesalahan yang dapat diterima dalam model dan mencari garis pemisah yang sesuai (atau dalam dimensi yang lebih tinggi disebut *hyperplane*) agar dapat melakukan proses regresi sesuai dengan data yang digunakan dalam pelatihan maupun pengujian model pembelajaran mesin. Sementara model regresi proses Gaussian (GPR) adalah model probabilistik berbasis

kernel non-parametrik. Kernel tersebut digunakan untuk memperkirakan fungsi kepadatan variabel acak, atau dalam regresi kernel untuk memperkirakan ekspektasi bersyarat dari variabel acak.

2. Model Pembelajaran Mesin untuk Masalah Regresi

2.1. SVR (support vector regression)

Metode SVM pertama kali diperkenalkan oleh Vladimir Vapnik bersama koleganya pada tahun 1992. SVM merupakan salah satu metode yang umum digunakan pada machine learning untuk masalah klasifikasi dan regresi. Khusus untuk regresi, SVM dipandang sebagai teknik nonparameter yang bergantung pada fungsi kernel. Pada makalah ini digunakan model Regresi SVM Linier dengan dual-konveks kuadratik dan Gaussian kernel. Misalkan kita memiliki satu set pasangan data latih $\langle x_n, y_n \rangle$, di mana x_n adalah himpunan multivarian dari N -observasi fitur input data latih, sementara y_n adalah target sampel data. Pada metode SVM, fungsi $f(x)$ akan dicari sebagai fungsi hyperplane (fungsi pemisah) yang memetakan vektor input ke dimensi yang lebih tinggi dengan eror sekecil mungkin. Fungsi regresi pemisah tersebut dapat ditulis sebagai berikut. Ilustrasi margin/garis pemisah maksimum diilustrasikan pada Gambar 2 dimana pada contoh ilustrasi tersebut menggunakan dua kelas pelatihan.

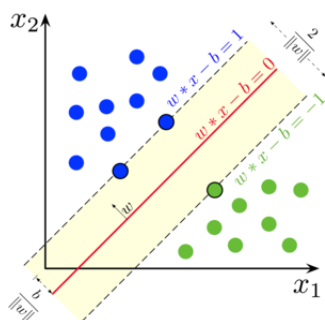
$$f(x) = w' \varphi(x) + b$$

dengan

w = vektor bobot,

$\varphi(x)$ = pemetaan fungsi ke dimensi yang lebih tinggi (fungsi kernel),

b = bobot bias.



Gambar 2. Hyperplane dan margin maksimum SVM menggunakan sampel pelatihan dua kelas, x_1 dan x_2 .

Selanjutnya didefinisikan fungsi objektif yang meminimasi vektor bobot berikut.

$$J(w) = \frac{1}{2} w' w$$

Pada regresi terdapat fungsi loss atau residual, yang menghitung ϵ yang menyatakan jarak antara fungsi estimasi $f(x)$ dengan output/target y yang dinyatakan dengan:

$$L_\epsilon = \begin{cases} 0 & , |y - f(x)| \leq \epsilon \\ |y - f(x)| - \epsilon & \text{selainnya} \end{cases}$$

Dalam bentuk umum dapat ditulis,

$$\forall n: |y_n - (w' \varphi(x_n) + b)| \leq \epsilon,$$

dimana n disini adalah data latih ke- n . Ada kemungkinan bahwa tidak ada fungsi $f(x)$ yang memenuhi batasan ini untuk semua titik. Untuk mengatasi batasan yang tidak memenuhi kondisi (infeasible) dan menghasilkan margin yang lebih halus, diperkenalkan variabel slack ξ_n dan ξ_n^* pada setiap titik, sehingga masalah optimasi yang baru menjadi berikut ini, dimana masalah optimasi berikut adalah masalah pemrograman dual-konveks [11], [12].

$$J(w) = \frac{1}{2} w' w + C \sum_{n=1}^N (\xi_n + \xi_n^*)$$

dengan batasan

$$\forall n: y_n - (w' \varphi(x_n) + b) \leq \epsilon + \xi_n,$$

$$\forall n: (w' \varphi(x_n) + b) - y_n \leq \varepsilon + \xi_n^*,$$

$$\forall n: \xi_n^* \geq 0,$$

$$\forall n: \xi_n \geq 0.$$

Di dalam makalah ini, menggunakan dua jenis kernel yaitu kernel kuadratik dan kernel Gaussian. Kernel kuadratik diberikan dengan fungsi berikut,

$$\varphi(x_j, x_k) = (1 + x_j' x_k)^2.$$

Sementara pada Kernel Gaussian memungkinkan variasi fungsi respons yang lebih cepat. Kernel Gaussian diberikan dengan fungsi berikut,

$$\varphi(x_j, x_k) = e^{-\|x_j - x_k\|^2}.$$

2.2. GPR (*Gaussian process regression*)

Diberikan himpunan data pelatihan berikut $\{(x_i, y_i); i = 1, 2, \dots, n\}$, yang diambil dari distribusi yang acak. Untuk proses regresi, model GPR akan dilatih untuk memprediksi nilai variabel respon y baru, diberikan masukan x baru. Model regresi linier diberikan sebagai berikut [13],

$$y = x^T \beta + \varepsilon,$$

dengan $\varepsilon \sim N(0, \sigma^2)$, dimana variansi kesalahan σ^2 dan koefisien bobot β diestimasi dari data. Proses Gaussian adalah himpunan dari variabel acak, sedemikian sehingga untuk bilangan terbatas (finite) memiliki distribusi Gaussian gabungan. Jika $f(x)$ adalah proses Gaussian, diberikan n -observasi x_1, \dots, x_n , maka distribusi gabungan dari variabel acak, $f(x_1), \dots, f(x_n)$, adalah Gaussian. Fungsi proses Gaussian ditentukan oleh fungsi rata-rata, $m(x)$, dan fungsi kovarian, $k(x, x')$. Jika $f(x)$ adalah proses Gaussian (GP), maka diperoleh ekspektasi rata-rata berikut,

$$E[f(x)] = m(x),$$

dan kovarian sebagai berikut,

$$\text{cov}[f(x), f(x')] = E[(f(x) - m(x))(f(x') - m(x')))] = k(x, x').$$

Model di atas dapat dituliskan sebagai berikut,

$$y = h(x)^T \beta + f(x),$$

dengan $f(x) \sim GP(0, k(x, x'))$,

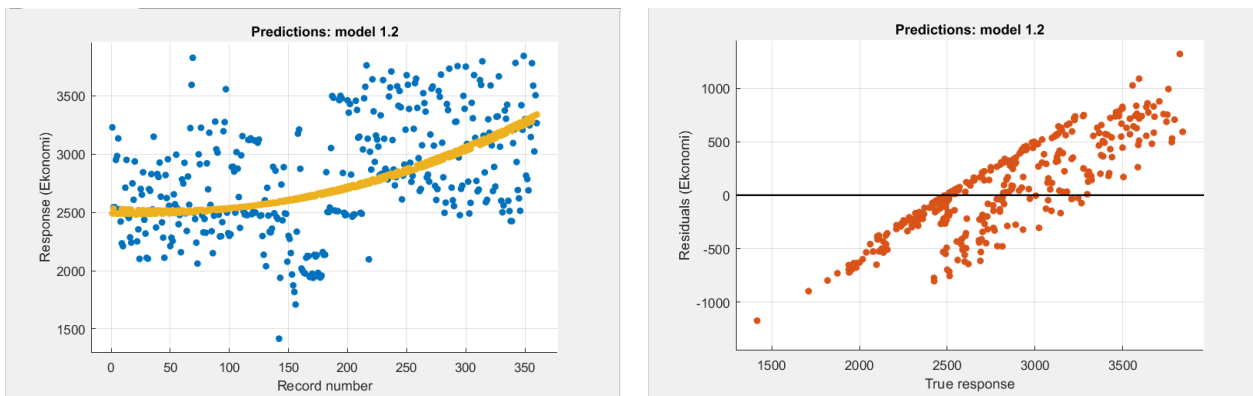
dimana rata-rata GP adalah nol dan kovarian diberikan oleh fungsi $k(x, x')$. $h(x)$ adalah himpunan dari fungsi basis yang memetakan fitur vektor x ke dalam fitur vektor baru $h(x)$. Berikut adalah contoh dari respon y dapat dimodelkan sebagai berikut,

$$P(y_i | f(x_i), x_i) \sim N(y_i | h(x_i)^T \beta + f(x_i), \sigma^2).$$

3. Regresi Okupansi Penumpang Kelas Ekonomi Kereta Api Argo Parahyangan (tahun 2019) dengan Metode SVR (*Support Vector Regression*)

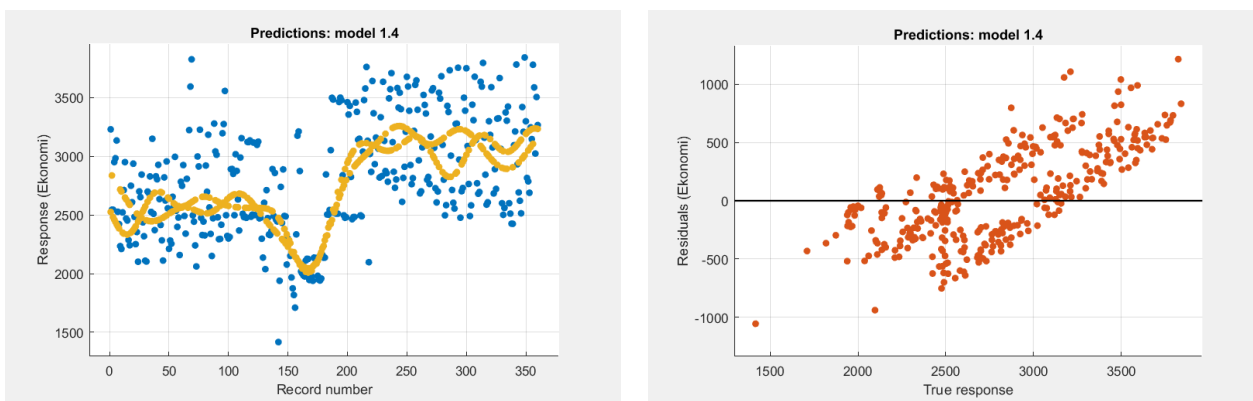
3.1. Simulasi regresi okupansi penumpang kereta api dengan metode SVR

Untuk kernel rasional kuadratik, yang diberikan oleh fungsi berikut $\varphi(x_j, x_k) = (1 + x_j' x_k)^2$, diperoleh hasil simulasi untuk prediksi penumpang kereta api seperti pada Gambar 3.



Gambar 3. Regresi okupansi penumpang kereta api dengan metode SVR (kernel kuadratik)

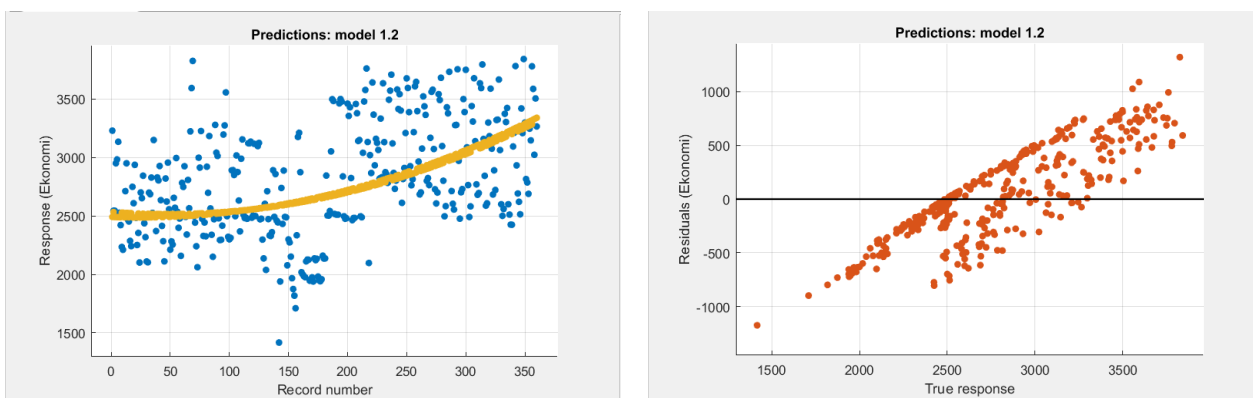
Untuk kernel fine Gaussian, yang diberikan oleh fungsi berikut $\varphi(x_j, x_k) = e^{-\|x_j - x_k\|^2}$, diperoleh hasil simulasi untuk prediksi penumpang kereta api seperti pada Gambar 4.



Gambar 4. Regresi okupansi penumpang kereta api dengan metode SVR (kernel eksponensial)

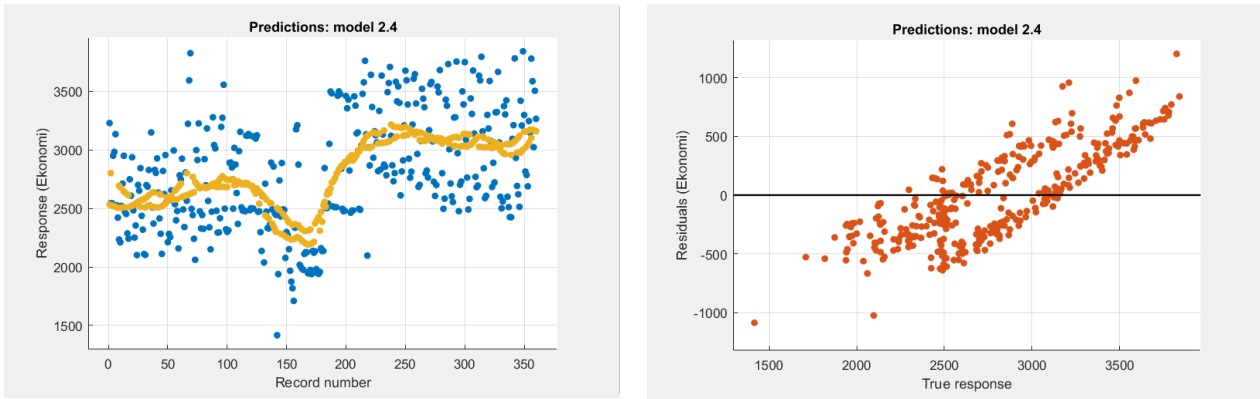
3.2. Simulasi regresi okupansi penumpang kereta api dengan metode GPR

Untuk kernel kuadratik, yang diberikan oleh fungsi berikut $h(x_j, x_k) = (1 + x_j'x_k)^2$, diperoleh hasil simulasi untuk prediksi penumpang kereta api seperti pada Gambar 5.



Gambar 5. Regresi okupansi penumpang kereta api dengan metode GPR (kernel kuadratik)

Untuk kernel eksponensial, yang diberikan oleh fungsi berikut $h(x_j, x_k) = e^{-\|x_j - x_k\|^2}$, diperoleh hasil simulasi untuk prediksi penumpang kereta api seperti pada Gambar 6.

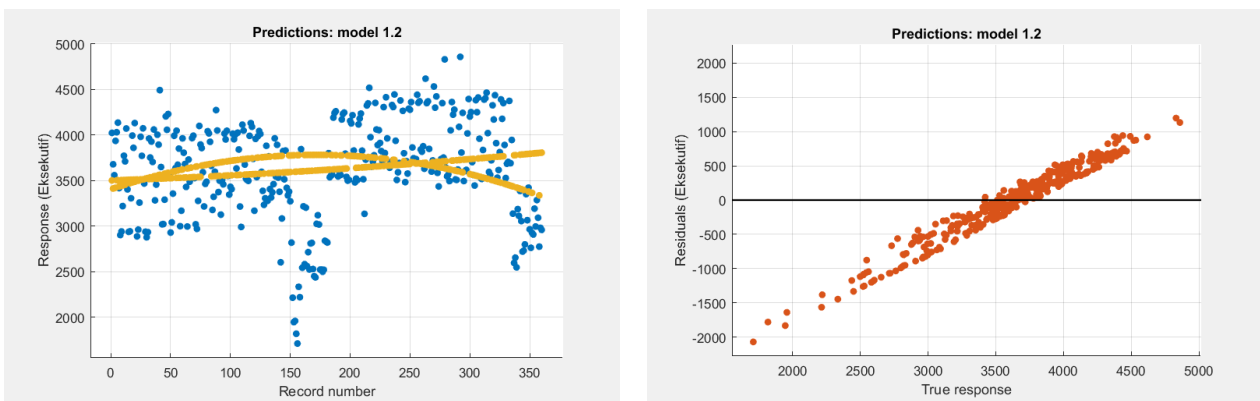


Gambar 6. Regresi okupansi penumpang kereta api dengan metode GPR (kernel eksponensial)

4. Regresi Okupansi Penumpang Kelas Eksekutif Kereta Api Argo Parahyangan (tahun 2019) dengan Metode GPR (Gaussian Process Regression)

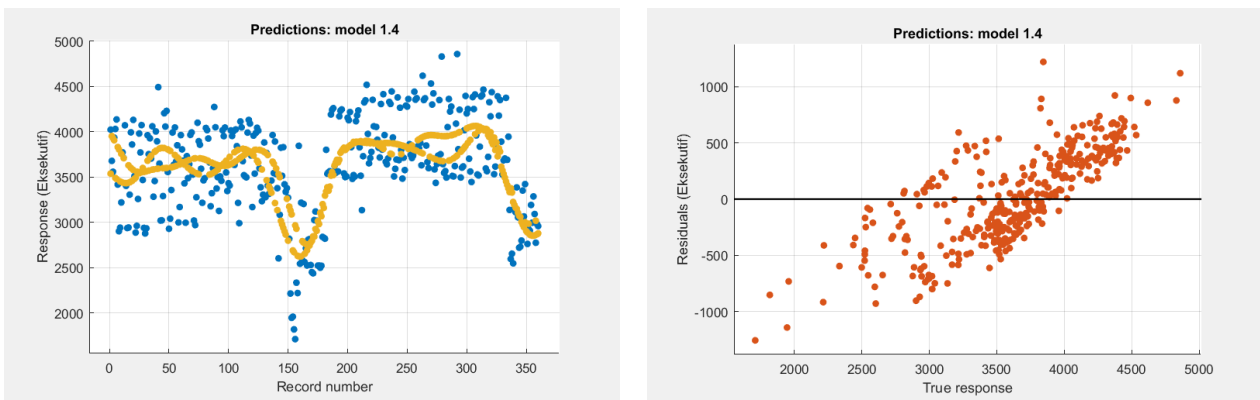
4.1. Simulasi regresi okupansi penumpang kereta api dengan metode GPR

Untuk kernel rasional kuadratik, yang diberikan oleh fungsi berikut $\varphi(x_j, x_k) = (1 + x_j'x_k)^2$, diperoleh hasil simulasi untuk prediksi penumpang kereta api seperti pada Gambar 7.



Gambar 7. Regresi okupansi penumpang kereta api dengan metode SVR (kernel kuadratik)

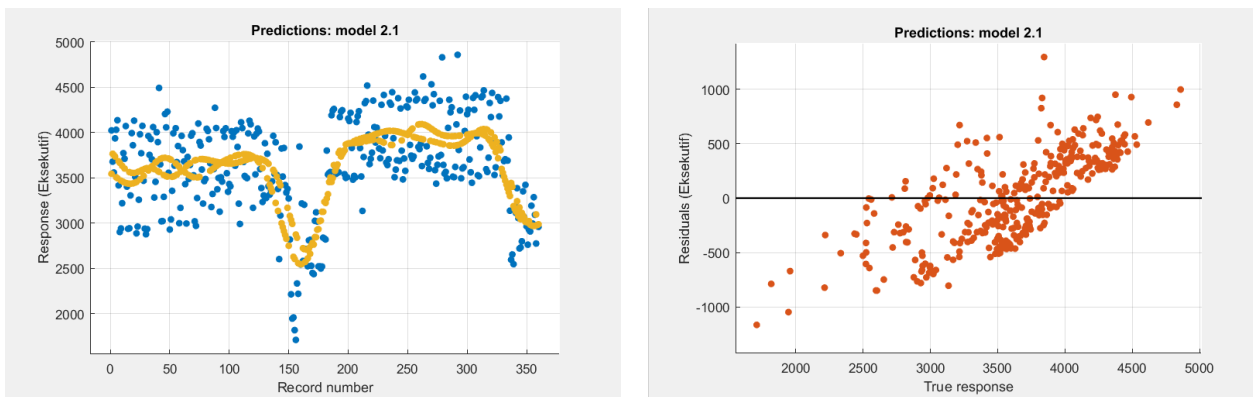
Untuk kernel fine Gaussian, yang diberikan oleh fungsi berikut $\varphi(x_j, x_k) = e^{-\|x_j - x_k\|^2}$, diperoleh hasil simulasi untuk prediksi penumpang kereta api seperti pada Gambar 8.



Gambar 8. Regresi okupansi penumpang kereta api dengan metode SVR (kernel eksponensial)

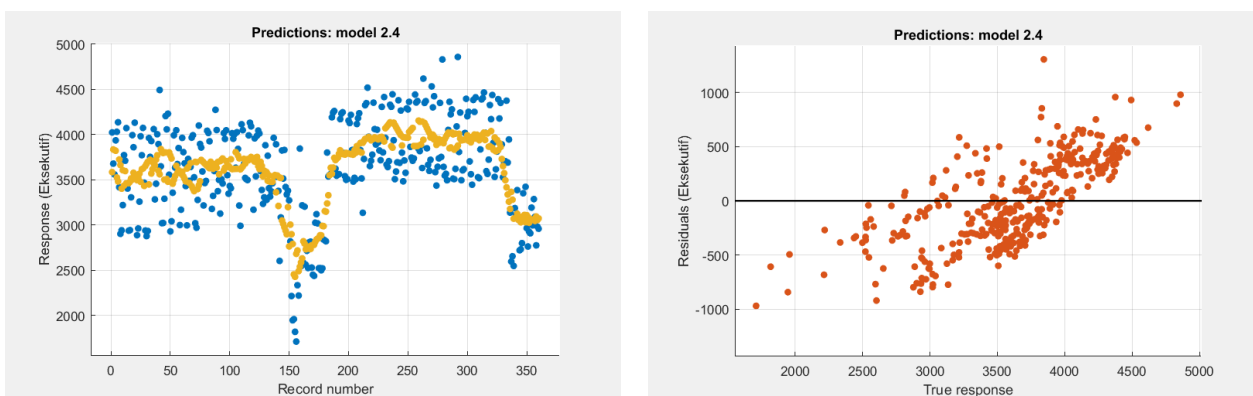
4.2. Simulasi regresi okupansi penumpang kereta api dengan metode GPR

Untuk kernel kuadratik, yang diberikan oleh fungsi berikut $h(x_j, x_k) = (1 + x_j'x_k)^2$, diperoleh hasil simulasi untuk prediksi penumpang kereta api seperti pada Gambar 9.



Gambar 9. Regresi okupansi penumpang kereta api dengan metode GPR (kernel kuadratik)

Untuk kernel eksponensial, yang diberikan oleh fungsi berikut $h(x_j, x_k) = e^{-\|x_j-x_k\|^2}$, diperoleh hasil simulasi untuk prediksi penumpang kereta api seperti pada Gambar 10.



Gambar 10. Regresi okupansi penumpang kereta api dengan metode GPR (kernel eksponensial)

Hasil komputasi beberapa parameter metriks evaluasi dari setiap model pembelajaran selanjutnya dievaluasi dengan sebagai berikut, dimana Tabel 1 adalah metriks evaluasi untuk regresi okupansi kereta api ekonomi dan Tabel 2 adalah metriks evaluasi untuk regresi okupansi kereta api eksekutif.

Tabel 1. Parameter dari beberapa metriks evaluasi regresi okupansi kereta ekonomi

Metriks Evaluasi	Quadratic SVR		Fine Gaussian SVR		Rational Quadratic GPR		Exponential GPR	
	2 folds	5 folds	2 folds	5 folds	2 folds	5 folds	2 folds	5 folds
RMSE	440.71	440.42	399.31	412.48	390.13	392.02	388.4	390.64
R-Squared	0.17	0.17	0.32	0.28	0.35	0.35	0.36	0.35
MAE	360.05	360.82	328.95	339.35	329.56	330.77	329.81	331.24
Kecepatan prediksi (obs/detik)	870	3100	110	4000	12000	7900	12000	7800
Waktu training (detik)	13.847	5.2427	11.33	4.8688	2.9395	5.7778	2.445	3.4752

Dari Tabel 1, untuk kinerja akurasi model prediksi untuk kereta api Argo Parahyangan kelas ekonomi diperoleh model terbaik berdasarkan metrik evaluasi di atas adalah model pembelajaran GPR (Gaussian process regression) dengan kernel eksponensial, dengan nilai parameter antara lain $RMSE = 388.4$, $R^2 = 0.36$, dan $MAE = 329.81$, dengan grafik regresi terlihat pada Gambar 5. Sementara untuk waktu komputasi tercepat juga dihasilkan dari metode GPR dengan kernel eksponensial, dengan waktu komputasi sebesar 2.445 detik, dengan grafik regresi terlihat pada Gambar 5, menggunakan prosesor Intel Core i5 generasi ke-8 dengan GPU AMD Radeon 535.

Tabel 2. Parameter dari beberapa metrik evaluasi regresi okupansi kereta eksekutif

Metriks Evaluasi	Quadratic SVR		Fine Gaussian SVR		Rational Quadratic GPR		Exponential GPR	
	2 folds	5 folds	2 folds	5 folds	2 folds	5 folds	2 folds	5 folds
RMSE	554.88	542.02	415.86	412.9	405.57	396.42	398.25	330.18
R-Squared	-0.05	0	0.41	0.42	0.44	0.46	0.46	0.63
MAE	433.53	422.8	343.53	345.44	340.36	336.87	337.5	258.24
Kecepatan prediksi (obs/detik)	1500	31000	1600	27000	13000	27000	9500	19000
Waktu training (detik)	8.5912	1.2612	8.1099	0.8767	3.7196	5.1933	2.8674	2.9669

Dari Tabel 2, untuk kinerja akurasi model prediksi untuk kereta api Argo Parahyangan kelas eksekutif diperoleh model terbaik berdasarkan metrik evaluasi di atas adalah model pembelajaran GPR dengan kernel eksponensial, dengan nilai parameter antara lain $RMSE = 330.18$, $R^2 = 0.63$, dan $MAE = 258.24$, dengan grafik regresi terlihat pada Gambar 10. Sementara untuk waktu komputasi tercepat dihasilkan dari metode pembelajaran SVR dengan kernel Gaussian, dengan waktu komputasi sebesar 2.445 detik, dengan grafik regresi terlihat pada Gambar 7, dan kecepatan prediksi per detik yaitu SVR dengan kernel kuadrat, dengan kecepatan prediksi sebesar 31000 prediksi/detik, dengan grafik regresi terlihat pada Gambar 8, menggunakan prosesor Intel Core i5 generasi ke-8 dengan GPU AMD Radeon 535.

5. Pengujian prediksi okupansi penumpang kereta api

Pengujian dilakukan dengan membandingkan data okupansi riil lapangan selama 5 hari ke depan dengan hasil prediksi okupansi penumpang, masing-masing menggunakan metode SVR dan GPR.

5.1. Prediksi okupansi penumpang kereta api untuk kelas ekonomi

Perbandingan prediksi okupansi penumpang kereta api dengan metode SVR dan GPR disajikan pada Tabel 3.

Tabel 3. Perbandingan prediksi okupansi penumpang kereta api ekonomi dengan data riil

Okupansi Kereta Ekonomi		SVR Quadratic	SVR Fine Gaussian	GPR Rational Quadratic	GPR Exponential
Data Prediksi (dalam orang)					
Hari	Data Real	2 folds	2 folds	2 folds	2 folds
361	2925	3322	3259	3156	3189
362	3177	3327	3263	3160	3183
363	3847	3332	3266	3164	3177
364	3476	3337	3267	3168	3171
365	2504	3342	3266	3170	3165
RMSE		483.094	464.9	460.1	457.9

Pada hasil pengujian untuk prediksi okupansi penumpang kereta api kelas ekonomi, secara keseluruhan metode pembelajaran GPR dalam pengujian ini memperoleh nilai RMSE yang lebih baik daripada metode SVR, dengan nilai RMSE okupansi total dari hasil prediksi dibandingkan data riil sebesar 457 penumpang dalam kurun 5 hari. Hasil pengujian dengan RMSE terkecil diperoleh menggunakan GPR dengan kernel eksponensial.

5.2. Prediksi okupansi penumpang kereta api untuk kelas eksekutif

Perbandingan prediksi okupansi penumpang kereta api dengan metode SVR dan GPR disajikan pada Tabel 4.

Tabel 4. Perbandingan prediksi okupansi penumpang kereta api eksekutif dengan data riil

Okupansi Kereta Eksekutif		SVR Quadratic	SVR Fine Gaussian	GPR Rational Quadratic	GPR Exponential
Data Prediksi (dalam orang)					
Hari	Data Real	2 folds	2 folds	2 folds	2 folds
361	2925	3613	3038	3073	3150
362	3177	3612	3056	3092	3279
363	3847	3610	3075	3113	3372
364	3476	3609	3095	3135	3439
365	2504	3608	3116	3158	3487
RMSE		625.339	478.1	471.6	500.9

Pada hasil pengujian untuk prediksi okupansi penumpang kereta api kelas eksekutif, secara keseluruhan metode pembelajaran GPR dalam pengujian ini memperoleh nilai RMSE yang lebih baik daripada metode SVR, dengan nilai RMSE okupansi total dari hasil prediksi dibandingkan data riil sebesar 471 penumpang dalam kurun 5 hari. Hasil pengujian dengan RMSE terkecil diperoleh menggunakan GPR dengan kernel kuadratik.

6. Kesimpulan

Pada proses pelatihan prediksi okupansi jumlah penumpang Kereta Api Argo Parahyangan pada tahun 2019 dengan menggunakan model pembelajaran SVR (*support vector regression*) dan model pembelajaran GPR (*Gaussian process regression*) diperoleh hasil kinerja terbaik untuk akurasi prediksi, baik kelas ekonomi maupun kelas eksekutif, yaitu menggunakan metode pembelajaran GPR dengan kernel eksponensial. Untuk waktu komputasi, untuk kelas ekonomi, proses komputasi tercepat yaitu menggunakan metode pembelajaran GPR dengan kernel eksponensial, dan untuk kelas eksekutif, proses komputasi tercepat yaitu menggunakan metode SVR dengan kernel Gaussian. Sementara untuk kecepatan prediksi paling tinggi, untuk kelas ekonomi yaitu menggunakan metode pembelajaran GPR dengan kernel eksponensial. Sementara untuk kelas eksekutif, kecepatan prediksi paling tinggi yaitu menggunakan metode pembelajaran SVR dengan kernel kuadratik.

Pada proses pengujian prediksi okupansi jumlah penumpang Kereta Api Argo Parahyangan pada tahun 2019 dengan menggunakan model pembelajaran SVR (*support vector regression*) dan model pembelajaran GPR (*Gaussian process regression*) untuk kurun 5 hari ke depan (pada akhir tahun 2019, mendekati tahun baru 2020). Pada kelas ekonomi, hasil kinerja terbaik untuk akurasi prediksi okupansi, yaitu menggunakan metode pembelajaran GPR dengan kernel eksponensial. Sementara pada kelas eksekutif, hasil kinerja terbaik untuk akurasi prediksi okupansi, yaitu menggunakan metode pembelajaran GPR dengan kernel kuadratik. Sehingga secara keseluruhan dapat disimpulkan, untuk kasus prediksi okupansi penumpang pada Kereta Api Argo Parahyangan, baik kelas ekonomi maupun kelas eksekutif, metode pembelajaran mesin untuk masalah regresi okupansi penumpang menggunakan metode GPR (*Gaussian process regression*) memiliki kinerja yang lebih baik dibandingkan menggunakan metode SVR (*support vector regression*).

Acknowledgment

Penulis mengucapkan terima kasih kepada Universitas Telkom atas dukungan terhadap penelitian ini. Penulis juga mengucapkan terima kasih kepada pihak PT. Kereta Api Indonesia (KAI) Daop II Bandung, atas bantuannya untuk memperoleh data jumlah penumpang Kereta Api Argo Parahyangan pada tahun 2019 untuk kelas ekonomi dan kelas eksekutif dengan narahubung mahasiswa Universitas Telkom atas nama Ufiq Nur Aeni dan Nur Asfinah.

References

- [1] Jamaludin, & Astuti, S. W. (2018). Kajian biaya operasi kereta api. *Jurnal Perkeretaapian Indonesia*, II(ISSN 2550-1127), 56–65.
- [2] Vandewiele, G., Colpaert, P., Janssens, O., Van Herwegen, J., Verborgh, R., Mannens, E., ... & De Turck, F. (2017). Predicting train occupancies based on query logs and external data sources. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 1469-1474).
- [3] Jenelius, E. (2019). Data-driven metro train crowding prediction based on real-time load data. *IEEE Transactions on Intelligent Transportation Systems*, 21(6), 2254-2265.
- [4] Guo, J., Xie, Z., Qin, Y., Jia, L., & Wang, Y. (2019). Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM. *IEEE Access*, 7, 42946-42955.
- [5] Guo, J., Xie, Z., Li, Q., Zhan, S., & Xu, J. (2020). Railway Passenger Flow Recognition Algorithm for Terminal Station Based on Cost Theory and Automatic Frequency Control. *IEEE Access*, 8, 26885-26892.
- [6] Wang, B., Ye, M., Zhu, Z., Li, Y., Liang, Q., & Zhang, J. (2020). Short-Term Passenger Flow Prediction for Urban Rail Stations Using Learning Network Based on Optimal Passenger Flow Information Input Algorithm. *IEEE Access*, 8, 170742-170753.
- [7] Jing, Z., & Yin, X. (2020). Neural Network-Based Prediction Model for Passenger Flow in a Large Passenger Station : An Exploratory Study. *IEEE Access*, 8, 36876–36884.
- [8] Lu, W., Ma, C., & Li, P. (2020). Research on Sample Selection of Urban Rail Transit Passenger Flow Forecasting Based on SCBP Algorithm. *IEEE Access*, 8, 89425-89438.
- [9] Lei, J., Dong, W., Chang, L., Xu, X., & Xu, X. (2019). Metro Traffic Prediction and Service-Oriented Scheduling Optimization Under Influence of Train Arrivals. *IEEE Access*, 7, 166729–166739.
- [10] Sun, Y., Cao, Y., & Roberts, C. (2019). A Hybrid Method for Life Prediction of Railway Relays Based on Multi-Layer Decomposition and RBFNN. *IEEE Access*, 7, 44761–44770.
- [11] Fan, R. E., Chen, P. H., & Lin, C. J. (2005). Working set selection using the second order information for training {SVM}. *Journal of Machine Learning Research*, 6, 1889–1918.
- [12] Chen, P. H., Fan, R. E., & Lin, C. J. (2006). A study on SMO-type decomposition methods for support vector machines. *IEEE Transactions on Neural Networks*, 17(4), 893–908.
- [13] Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. In MIT Press. ISSN: 10236090.