

HUBUNGAN DATA MELALUI ANALISIS REGRESI DALAM PENDEKATAN VISUALISASI: KAJIAN LITERATUR

Septia Ayu Pratiwi¹, Wicaksono Yuli Sulisty²

^{1,2}Sistem Informasi, Fakultas Teknologi dan Kesehatan, Universitas Siber Muhammadiyah, Yogyakarta

Abstrak | Dalam era *big data*, analisis regresi menjadi alat yang sangat penting untuk memahami dan menginterpretasikan hubungan antar variabel. Pendekatan visualisasi data dalam analisis regresi tidak hanya memperkuat pemahaman tentang hubungan ini tetapi juga memudahkan pengguna untuk mengeksplorasi dan mengkomunikasikan temuan. Artikel ini bertujuan untuk mengeksplorasi berbagai teknik visualisasi yang dapat digunakan dalam analisis regresi, dari plot sederhana hingga visualisasi yang lebih kompleks, untuk membantu dalam interpretasi model regresi. Kami juga membahas bagaimana visualisasi dapat digunakan untuk diagnostik model dan asumsi regresi, serta untuk menunjukkan hubungan non-linear atau interaksi antar variabel.

Kata Kunci: Analisis Regresi, Diagnostik Model, Hubungan Non-Linear, Model Regresi, Visualisasi Data

Abstract | Di era analisis big data, regresi menjadi alat yang sangat penting untuk memahami dan menginterpretasikan hubungan antar variabel. Pendekatan visualisasi data analisis dalam regresi tidak hanya memperkuat pemahaman tentang hubungan ini tetapi juga memudahkan pengguna untuk mengeksplorasi dan mengkomunikasikan temuan. Artikel ini bertujuan untuk mengeksplorasi berbagai teknik visualisasi yang dapat digunakan dalam analisis regresi, dari plot sederhana hingga visualisasi yang lebih kompleks, untuk membantu dalam interpretasi model regresi. Kami juga membahas bagaimana visualisasi dapat digunakan untuk diagnostik model dan asumsi regresi, serta untuk menunjukkan hubungan non-linear atau interaksi antar variabel.

Keywords: *Regression Analysis, Model Diagnostics, Non-Linear Relationships, Regression Models, Data Visualization*

Pendahuluan

Analisis regresi merupakan salah satu metodologi statistik paling fundamental dan sering digunakan dalam berbagai disiplin ilmu untuk memahami hubungan antara variabel. Sejak diperkenalkan, teknik ini telah berkembang dan menyesuaikan dengan kebutuhan analitis yang beragam, mengakomodasi data yang lebih kompleks dan pertanyaan penelitian yang lebih sofistikasi. Dalam dekade terakhir, pertumbuhan eksponensial dalam volume data yang tersedia untuk analisis telah mendorong pengembangan teknik analitis yang lebih canggih, termasuk dalam analisis regresi. Visualisasi data, sebagai pendekatan untuk menginterpretasikan data kompleks secara intuitif, telah menjadi kunci dalam mengkomunikasikan temuan dan insight dari model regresi.

Banyak penelitian telah menunjukkan pentingnya visualisasi dalam analisis regresi. Menurut Grömping (2015), visualisasi dapat secara signifikan meningkatkan pemahaman tentang struktur data dan hubungan antar variabel. Selanjutnya, Breheny & Burchett (2017) menekankan bahwa visualisasi data memungkinkan peneliti untuk mengidentifikasi pola, tren, dan outlier yang mungkin terlewat oleh analisis statistik tradisional. Dalam konteks saat ini, teknologi visualisasi data telah berkembang pesat, memberikan alat yang lebih dinamis dan interaktif untuk eksplorasi data.

Kajian terbaru oleh Avraam et al. (2021) menggarisbawahi pentingnya integrasi teknik visualisasi dalam analisis regresi untuk membuat model lebih dapat diakses dan dimengerti oleh pengguna non-teknis. Hal ini menjadi semakin relevan di era *big data*, di mana kemampuan untuk dengan cepat menginterpretasikan dan mengambil keputusan berdasarkan data menjadi kritis. Selain itu, penelitian oleh (Kang et al., 2020). Kaur & Kaur (2020) menunjukkan bahwa visualisasi dapat memainkan peran vital dalam diagnostik model regresi, membantu dalam identifikasi masalah seperti heteroskedastisitas, multicollinearity, dan non-linearitas yang mungkin mempengaruhi keakuratan model.

Pendekatan ini tidak hanya meningkatkan transparansi dalam proses analitis tetapi juga memfasilitasi kolaborasi multidisiplin, memungkinkan individu dari berbagai latar belakang untuk berkontribusi pada interpretasi dan pengambilan keputusan berdasarkan data. Oleh karena itu, integrasi antara analisis regresi dan visualisasi data bukan hanya merupakan perkembangan metodologis tetapi juga evolusi dalam cara kita memahami dan berinteraksi dengan data.

Metode

Metodologi yang diadopsi dalam studi ini bertujuan untuk mengevaluasi efektivitas berbagai teknik visualisasi dalam memfasilitasi pemahaman dan interpretasi model analisis regresi melalui kajian literatur. Pendekatan ini dibagi menjadi beberapa tahap utama, yaitu pemilihan data, analisis regresi, penerapan teknik visualisasi, dan evaluasi efektivitas visualisasi. Kajian ini berusaha untuk mengidentifikasi bagaimana visualisasi data dapat meningkatkan pengertian tentang hubungan antar variabel dan efektivitas diagnostik model regresi.

Tujuan Penelitian

Tujuan dari kajian literatur ini adalah untuk menyelidiki dan menganalisis penggunaan analisis regresi dalam konteks visualisasi data, dengan fokus pada perkembangan terbaru dalam metodologi, teknologi, aplikasi praktis, dan tantangan yang dihadapi. Kajian ini bertujuan untuk mengidentifikasi tren utama, gap dalam literatur yang ada, dan arah untuk penelitian selanjutnya.

Sumber Data

Kajian ini akan menggunakan sumber-sumber primer dan sekunder, termasuk artikel jurnal ilmiah, konferensi, tesis, disertasi, laporan teknis, dan buku. Database yang akan digunakan meliputi IEEE Xplore, PubMed, Scopus, Web of Science, dan Google Scholar untuk menggali kedalaman penelitian tentang kajian analisis regresi dalam visualisasi data. Artikel yang digunakan dalam penelitian ini diterbitkan antara tahun 2015 sampai dengan 2023 yang berfokus pada analisis regresi dan visualisasi data serta mencakup studi empiris, teoritis, atau review.

Strategi Pencarian

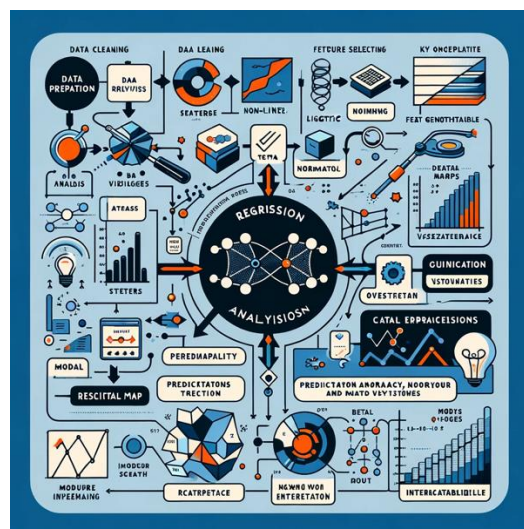
Kata kunci yang akan digunakan dalam pencarian termasuk "analisis regresi", "visualisasi data", "regresi linier", "regresi non-linier", "regresi logistik", "software visualisasi data", "AR dalam visualisasi data", "VR dalam visualisasi data", dan "etika dalam visualisasi data". Strategi pencarian akan disesuaikan untuk setiap database untuk memaksimalkan relevansi dan cakupan.

Seleksi dan Evaluasi Sumber

Setelah pengumpulan data awal, sumber akan dievaluasi berdasarkan relevansi, kebaruan, dan kualitas metodologis. Proses seleksi akan menggunakan metode review sistematis, termasuk penilaian independen oleh dua peneliti untuk memastikan objektivitas dan mengurangi bias.

Analisis Data

Adapun analisis data dilakukan dengan mengidentifikasi trend penggunaan analisis regresi dalam visualisasi data, gap dalam literatur dan tantangan yang dihadapi dalam penggunaan analisis regresi. Selain itu, studi literatur ini mengkaji perkembangan metodologis, aplikasi dan studi kasus penggunaan uji regresi serta kontribusinya dalam penelitian selanjutnya. Analisis ini akan dilakukan melalui sintesis kualitatif, dengan mengkategorikan temuan ke dalam tema utama sesuai dengan peta konsep yang telah dikembangkan. Adapun peta konsep terdapat pada Gambar 1.



Gambar 1. Peta Konsep Analisis Regresi untuk Visualisasi Data

Peta konsep difokuskan pada penggunaan uji regresi yaitu regresi linear, regresi non-linear, dan regresi logistik sebagai respons kuantitatif untuk menguji hubungan antar variabel. Selanjutnya

persiapan data digunakan untuk proses pembersihan data, seleksi variabel yang relevan dalam model, dan normalisasi untuk memastikan tidak ada variabel yang mendominasi. Teknik visualisasi menggunakan scatter plot, grafik garis, dan HeatMap untuk merepresentasikan visual dari hubungan dua variabel. Overview merupakan step akhir untuk menemukan kompleksitas data dan interpretasi data untuk menemukan wawasan baru dari data yang telah direpresentasikan.

Validasi Temuan

Temuan dari kajian literatur akan divalidasi melalui konsultasi dengan ahli di bidang analisis regresi dan visualisasi data berupa peer-review dengan ahli bidang analisis dan visualisasi data untuk memastikan keakuratan dan relevansi temuan penelitian. Hasil kajian kemudian disintesis menjadi sebuah laporan kajian literatur yang mencakup implikasi teoritis dan praktis serta rekomendasi untuk penelitian selanjutnya.

Hasil dan Pembahasan

Pengumpulan data

Data yang digunakan dalam kajian literatur ini berasal dari tahun 2015 sampai dengan 2023 dengan fokus penggunaan uji regresi dan visualisasi data. Untuk mendukung kedua temuan tersebut, kajian yang mendukung seperti aplikasi analisis regresi untuk representasi data di berbagai bidang, seperti bidang kesehatan, ekonomi, dan pendidikan juga dipertimbangkan sebagai pembahasan utama. Adapun kajian literatur diperoleh dengan mengkategorikan beberapa aspek, yaitu kajian utama, tren, aplikasi dan rekomendasi untuk kajian selanjutnya, seperti dijelaskan dalam Tabel 1.

Tabel 1. Kajian Analisis Regresi dan Visualisasi Data

Kategori	Deksripsi	Referensi
Kajian utama	Kemajuan dalam teknologi imersif yang meningkatkan keterlibatan pengguna melalui umpan balik sensorik multimodal dan teknik interaksi	A survey of immersive visualization: Focus on perception and interaction (Zhang et al., 2023)
	Identifikasi dan perbandingan berbagai pendekatan untuk mengoptimalkan akurasi dan presisi prediksi dalam model regresi linier dan polinomial, mengoptimalkannya berdasarkan karakteristik kumpulan data untuk mencapai kinerja yang lebih baik.	A Review on Linear Regression Comprehensive in Machine Learning (Maulud & Abdulazeez, 2020)
	Aplikasi regresi linear dalam multidisiplin dalam scope kesehatan publik dan keuangan	Correlation and simple linear regression (Zou et al., 2003, Marrie et al., 2009, Olsen et al., 2020)
Tren	Peningkatan interaktivitas dan penggunaan visualisasi data sebagai alat storytelling	Zhang, Y., Reynolds, M., Lugmayr, A., Damjanov, K., & Hassan, G. (2022). A Visual Data Storytelling Framework (Zhang et al., 2022)

	Pengembangan alat visualisasi berbasis web untuk akses dan kolaborasi yang lebih luas	Applications of the Web-based collaborative visualization in distributed product development (Chu et al., 2006)
Aplikasi	Visualisasi penyebaran COVID-19 dan prediksi kebutuhan sumber daya medis	Forecasting the novel coronavirus COVID-19 (Petropoulos & Makridakis, 2020)
	Penggunaan analisis regresi visual dalam prediksi pergerakan pasar dan resiko investasi	Predicting the price of a stock using regression analysis (Rudzeyt et al., 2020)

Analisis Regresi untuk Visualisasi Data

Teknik visualisasi data telah berkembang secara signifikan untuk mengatasi tantangan dalam menginterpretasikan *dataset* yang semakin besar dan kompleks. Teknik visualisasi data tidak hanya membantu dalam pemahaman dan analisis data tetapi juga memainkan peran krusial dalam proses pengambilan keputusan. Mulai dari visualisasi interaktif hingga pengembangan algoritma baru. Dalam hal ini, visualisasi interaktif dari *Big Data* menekankan pentingnya penggunaan visualisasi yang mudah untuk dipahami dan fleksibel untuk analisis berdasarkan data praktis, serta menemukan solusi dari permasalahan ukuran dataset. Visualisasi interaktif membawa solusi inovatif untuk pemrosesan query dan visualisasi data untuk meningkatkan pemahaman pengguna terhadap *Big Data set* (Godfrey et al., 2016). Penting juga menekankan teknik dan alat visualisasi data yang menyoroti signifikansi dalam menganalisis dan memahami *Big Dataset*. Ulasan ini menekankan peran visualisasi dalam mengidentifikasi outlier, tren, dan pola, yang krusial untuk pengambilan keputusan yang terinformasi (Kaur & Kaur, 2020). Pemrograman Genetik memperkenalkan pendekatan pemrograman genetik untuk mengembangkan pemetaan *interpretable* dari dataset ke visualisasi. Metode ini bertujuan untuk menyeimbangkan kualitas visual dengan kompleksitas model, sehingga data lebih mudah diterjemahkan. Berfokus pada analisis data eksploratif dan interaktif, teknik visualisasi data penting untuk diterapkan dalam alur kerja analitik. Selain untuk mendukung kemudahan merepresentasikan data, teknik visualisasi data dapat digunakan untuk meninjau kesenjangan antara penelitian tentang visualisasi data dengan teknik yang diterapkan untuk mengolah data (Schmidt, 2020).

Seiring berjalannya waktu, penelitian dalam analisis regresi telah menunjukkan perkembangan signifikan di berbagai bidang. Berdasarkan trend visualisasi data dan implikasinya bagi analisis regresi, (Hefley et al., 2017) menggunakan pendekatan fungsi basis untuk mengatasi autokorelasi dalam data ekologi, memperbaiki model untuk menggambarkan data yang terkait secara spasial atau temporal dengan lebih akurat. Stulp & Sigaud (2015) mengintegrasikan berbagai algoritma regresi dalam satu model terpadu, memfasilitasi pengembangan metode baru yang lebih efisien. Smith (2015) (Smith, 2015) menekankan pentingnya pemilihan model yang tepat dan interpretasi hasil yang cermat dalam analisis regresi. Kowal & Bourgeois (2020) mengembangkan kerangka kerja Bayesian untuk regresi fungsi terhadap skalar, menggunakan fungsi basis yang tidak diketahui untuk menciptakan basis fungsional yang fleksibel dan adaptif. Aviral Gupta et al. (2017) meninjau berbagai model regresi yang ada, dari Regresi Linier hingga Regresi Ridge dan Lasso, menekankan pentingnya pemilihan model yang sesuai dengan jenis data.

Kesimpulan

Analisis regresi dalam pendekatan visualisasi dalam kajian literatur menyoroti bagaimana visualisasi dapat meningkatkan pemahaman dan interpretasi hubungan antara variabel dalam kumpulan data. Hubungan ini biasanya dieksplorasi melalui model regresi, yang bertujuan untuk memodelkan hubungan antara variabel dependen (atau variabel target) dengan satu atau lebih variabel independen (atau prediktor).

Pendekatan visualisasi memungkinkan peneliti dan analis untuk memvisualisasikan hubungan tersebut secara intuitif, membuat asumsi model lebih mudah dipahami, dan mengidentifikasi potensi masalah dalam data atau model, seperti heteroskedastisitas, multikolinearitas, atau outlier. Visualisasi seperti plot scatter, diagram residu, dan kurva regresi dapat membantu dalam menilai kecocokan model dan kekuatan serta bentuk hubungan antar variabel.

Secara keseluruhan, pendekatan visualisasi dalam analisis regresi memperkaya proses analitis dengan menyediakan cara yang lebih dinamis dan interaktif untuk menganalisis dan menyajikan data. Ini tidak hanya meningkatkan pemahaman analitis tentang hubungan dalam data tetapi juga membantu dalam komunikasi temuan penelitian yang efektif kepada audiens yang lebih luas, termasuk mereka yang mungkin tidak memiliki latar belakang statistik yang kuat. Kajian literatur menggarisbawahi pentingnya integrasi visualisasi dalam analisis regresi sebagai alat vital untuk eksplorasi data, validasi model, dan pembuatan keputusan berbasis bukti.

Daftar Referensi

- Aviral Gupta, Akshay Sharma, & Dr. Amita Goel. (2017). Review of Regression Analysis Models. *International Journal of Engineering Research And*, V6(08). <https://doi.org/10.17577/IJERTV6IS080060>
- Avraam, D., Wilson, R., Butters, O., Burton, T., Nicolaidis, C., Jones, E., Boyd, A., & Burton, P. (2021). Privacy preserving data visualizations. *EPJ Data Science*, 10(1), 2. <https://doi.org/10.1140/epjds/s13688-020-00257-4>
- Breheny, P., & Burchett, W. (2017). Visualization of Regression Models Using visreg. *The R Journal*, 9(2), 56. <https://doi.org/10.32614/RJ-2017-046>
- Chu, C.-H., Cheng, C.-Y., & Wu, C.-W. (2006). Applications of the Web-based collaborative visualization in distributed product development. *Computers in Industry*, 57(3), 272–282. <https://doi.org/10.1016/j.compind.2005.12.004>
- Godfrey, P., Gryz, J., & Lasek, P. (2016). Interactive Visualization of Large Data Sets. *IEEE Transactions on Knowledge and Data Engineering*, 28(8), 2142–2157. <https://doi.org/10.1109/TKDE.2016.2557324>
- Grömping, U. (2015). Variable importance in regression models. *WIREs Computational Statistics*, 7(2), 137–152. <https://doi.org/10.1002/wics.1346>
- Hefley, T. J., Broms, K. M., Brost, B. M., Buderman, F. E., Kay, S. L., Scharf, H. R., Tipton, J. R., Williams, P. J., & Hooten, M. B. (2017). The basis function approach for modeling autocorrelation in ecological data. *Ecology*, 98(3), 632–646. <https://doi.org/10.1002/ecy.1674>
- Kang, X., Chen, X., Jin, R., Wu, H., & Deng, X. (2020). Multivariate regression of mixed responses for evaluation of visualization designs. *IIEE Transactions*, 53(3), 313–325. <https://doi.org/10.1080/24725854.2020.1755068>

- Kaur, R., & Kaur, B. (2020). DATA VISUALIZATION TECHNIQUES: A REVIEW. *Global Journal of Engineering Science and Researches*, 7(8), 18–26. <https://doi.org/10.29121/gjesr.v7.i8.2020.3>
- Kowal, D. R., & Bourgeois, D. C. (2020). Bayesian Function-on-Scalars Regression for High-Dimensional Data. *Journal of Computational and Graphical Statistics*, 29(3), 629–638. <https://doi.org/10.1080/10618600.2019.1710837>
- Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1(2), 140–147. <https://doi.org/10.38094/jastt1457>
- Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLOS ONE*, 15(3), e0231236. <https://doi.org/10.1371/journal.pone.0231236>
- Rudzeyt, O., Zainetdinov, A., Nedyak, A., & Ragulin, P. (2020). Predicting the price of a stock using regression analysis. *Russian Journal of Resources, Conservation and Recycling*, 7(4). <https://doi.org/10.15862/14INOR420>
- Schmidt, J. (2020). Usage of Visualization Techniques in Data Science Workflows. *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 309–316. <https://doi.org/10.5220/0009181903090316>
- Smith, G. (2015). The Art of Regression Analysis. In *Essential Statistics, Regression, and Econometrics* (pp. 261–299). Elsevier. <https://doi.org/10.1016/B978-0-12-803459-0.00009-1>
- Stulp, F., & Sigaud, O. (2015). Many regression algorithms, one unified model: A review. *Neural Networks*, 69, 60–79. <https://doi.org/10.1016/j.neunet.2015.05.005>
- Zhang, Y., Reynolds, M., Lugmayr, A., Damjanov, K., & Hassan, G. M. (2022). A Visual Data Storytelling Framework. *Informatics*, 9(4), 73. <https://doi.org/10.3390/informatics9040073>
- Zhang, Y., Wang, Z., Zhang, J., Shan, G., & Tian, D. (2023). A survey of immersive visualization: Focus on perception and interaction. *Visual Informatics*, 7(4), 22–35. <https://doi.org/10.1016/j.visinf.2023.10.003>
- Zou, K. H., Tuncali, K., & Silverman, S. G. (2003). Correlation and Simple Linear Regression. *Radiology*, 227(3), 617–628. <https://doi.org/10.1148/radiol.2273011499>