

Analisis Komparatif Algoritma Ridge, LASSO, MLP, dan Gaussian Process Regressor pada Prediksi Efisiensi Kendaraan Listrik

Annida Purnamawati^{1*}, Monikka Nur Winnarto², Mely Mailasari³

¹ Program Studi Sistem Informasi Kampus Kota Yogyakarta, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
Jl. Ringroad Barat, Amberketawang, Gamping, Sleman

² Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98 RT/2 RW/6 Kwitang, Kec. Senen, Jakarta Pusat, DKI Jakarta
Email: ^{1,*}annida.npr@bsi.ac.id, ²monikka.mnt@bsi.ac.id, ³mely.mly@bsi.ac.id,
Email Penulis Korespondensi: annida.npr@bsi.ac.id

Abstrak—Fenomena *range anxiety* pada pengguna kendaraan listrik menuntut sistem prediksi efisiensi yang akurat dan adaptif. Penelitian ini menawarkan kebaruan dengan analisis komparatif empat algoritma regresi: Ridge Regression, LASSO Regression, Multi-Layer Perceptron (MLP), dan Gaussian Process Regressor (GPR), menggunakan Electric Vehicle Population Data. Pendekatan CRISP-DM diterapkan secara sistematis, meliputi pemahaman bisnis, eksplorasi data, seleksi fitur berbasis korelasi dan multikolinearitas, pemodelan, serta evaluasi dengan metrik MAE, RMSE, dan R^2 melalui k-fold cross-validation. Hasil penelitian menunjukkan bahwa LASSO Regression unggul dalam akurasi rata-rata dan seleksi fitur dengan R^2 tertinggi (0.974), meskipun disparitas antara MAE (2.13) dan RMSE (16.10) mengindikasikan sensitivitas terhadap outlier. Ridge Regression dan GPR memberikan hasil lebih stabil dengan R^2 sekitar 0.93 dan error seimbang, menjadikannya pilihan konsisten untuk data kompleks. Sementara itu, MLP mampu menangkap pola non-linear, tetapi gagal memodelkan variabel teknis dominan seperti Range Efficiency dan Vehicle Age, sehingga menghasilkan error lebih tinggi (R^2 0.729). Novelty penelitian ini terletak pada integrasi analisis korelasi fitur dengan evaluasi komparatif empat algoritma regresi berbeda karakteristik matematisnya, sehingga tidak hanya menilai akurasi, tetapi juga kestabilan, interpretabilitas, dan sensitivitas terhadap outlier. Temuan ini memberikan insight bagi industri otomotif dan energi bahwa efisiensi kendaraan listrik sangat dipengaruhi faktor teknis seperti tahun produksi, usia kendaraan, dan kapasitas baterai. Model linear dengan regularisasi terbukti lebih efektif dalam menangkap pengaruh langsung variabel teknis. Selain itu, eksplorasi algoritma ensemble direkomendasikan untuk meningkatkan akurasi dan ketahanan model.

Kata kunci : Ridge Regression; LASSO Regression; Multi-Layer Perceptron; Gaussian Process Regressor; CRISP-DM

Abstract—The phenomenon of range anxiety among electric vehicle (EV) users necessitates the development of accurate and adaptive efficiency prediction systems. This study introduces novelty by conducting a comparative analysis of four regression algorithms Ridge Regression, LASSO Regression, Multi-Layer Perceptron (MLP), and Gaussian Process Regressor (GPR) using the Electric Vehicle Population Data. The CRISP-DM methodology was systematically applied, encompassing business understanding, data exploration, feature selection based on correlation and multicollinearity, modeling, and evaluation with MAE, RMSE, and R^2 metrics through k-fold cross-validation. The results indicate that LASSO Regression excels in average accuracy and feature selection, achieving the highest R^2 (0.974), although the disparity between MAE (2.13) and RMSE (16.10) highlights sensitivity to outliers. Ridge Regression and GPR deliver more stable outcomes with R^2 values around 0.93 and balanced error metrics, making them consistent choices for complex data distributions. In contrast, MLP captures non-linear patterns but fails to effectively model dominant technical variables such as Range Efficiency and Vehicle Age, resulting in higher error (R^2 0.729). The novelty of this research lies in integrating feature correlation analysis with a comparative evaluation of four regression algorithms with distinct mathematical characteristics, thereby assessing not only accuracy but also stability, interpretability, and robustness to outliers. The findings provide critical insights for the automotive and energy sectors, emphasizing that EV efficiency is strongly influenced by technical factors such as production year, vehicle age, and battery capacity. Linear models with regularization prove more effective in capturing direct impacts of these variables, while ensemble approaches are recommended to further enhance accuracy and resilience.

Keywords: Ridge Regression; LASSO Regression; Multi-Layer Perceptron; Gaussian Process Regressor; CRISP-DM

1. PENDAHULUAN

Kendaraan listrik (Electric Vehicle/EV) merupakan salah satu inovasi transportasi masa depan yang dinilai lebih ramah lingkungan dan berkelanjutan dibandingkan kendaraan berbahan bakar fosil[1]. Keunggulan utama kendaraan listrik terletak pada emisi karbon yang lebih rendah, sehingga mampu mengurangi dampak negatif terhadap lingkungan, khususnya polusi udara dan pemanasan global. Selain itu, penggunaan kendaraan listrik juga mendukung transisi menuju sistem energi bersih melalui pemanfaatan sumber energi terbarukan sebagai pengganti bahan bakar minyak[2]. Seiring dengan meningkatnya kesadaran masyarakat terhadap isu perubahan iklim dan keterbatasan sumber daya energi fosil, adopsi kendaraan listrik terus mengalami peningkatan secara global dari tahun ke tahun. Meskipun demikian, pengembangan kendaraan listrik masih menghadapi berbagai tantangan, salah satunya adalah efisiensi energi. Efisiensi kendaraan listrik menjadi faktor penting karena berkaitan langsung dengan konsumsi energi, jarak tempuh, serta performa kendaraan secara keseluruhan. Ketidakmampuan dalam memperkirakan efisiensi secara akurat dapat menyebabkan ketidakefisienan penggunaan energi dan menurunkan kepercayaan pengguna terhadap kendaraan listrik. Hal ini juga berpengaruh terhadap pengambilan keputusan dalam pengelolaan energi dan perencanaan penggunaan kendaraan.

Salah satu permasalahan utama dalam penggunaan kendaraan listrik adalah fenomena *range anxiety*, yaitu rasa khawatir pengguna bahwa daya baterai tidak mencukupi untuk mencapai tujuan perjalanan[3]. Kondisi ini menjadi isu psikologis yang cukup dominan di kalangan pengguna kendaraan listrik, khususnya bagi mereka yang sering melakukan perjalanan jarak jauh atau berada di daerah dengan ketersediaan stasiun pengisian daya yang terbatas. Kekhawatiran tersebut tidak hanya memengaruhi tingkat kenyamanan dalam berkendara, tetapi juga berdampak pada kepercayaan

pengguna terhadap teknologi kendaraan listrik. Lebih jauh lagi, kondisi ini turut menjadi hambatan dalam mendorong adopsi kendaraan listrik secara luas di masyarakat. Banyak yang menilai kendaraan listrik belum cukup fleksibel, terutama dari sisi jarak tempuh dan ketersediaan energi[4]. Kekhawatiran ini umumnya muncul karena estimasi sisa jarak tempuh yang kurang akurat. Meskipun informasi tersebut ditampilkan melalui sistem onboard kendaraan, dalam praktiknya data yang diberikan sering kali belum cukup presisi dan tidak selalu diperbarui secara real-time.

Untuk mengatasi permasalahan *range anxiety* yang disebabkan oleh estimasi jarak tempuh yang kurang akurat, pendekatan berbasis *Data Science* dan algoritma *Machine Learning* (ML) dinilai sebagai solusi yang potensial[5]. Pendekatan ini memungkinkan pemanfaatan data historis penggunaan kendaraan listrik secara lebih mendalam, sehingga model prediksi yang dihasilkan menjadi lebih akurat dan adaptif. Dengan demikian, estimasi jarak tempuh kendaraan dapat dilakukan secara lebih dinamis dan mendekati kondisi nyata (*real-time*)[6]. Algoritma regresi menjadi metode yang umum digunakan karena mampu memodelkan hubungan antara variabel input dan output. Beberapa algoritma yang sering digunakan antara lain Ridge Regression dan Least Absolute Shrinkage and Selection Operator (LASSO) yang termasuk dalam regresi regularisasi, Multi-Layer Perceptron (MLP) berbasis jaringan saraf tiruan, serta Gaussian Process Regressor yang bersifat probabilistik. Dengan memanfaatkan data seperti Tahun, Model, Merek, Jenis Kendaraan Listrik, Kendaraan Bahan Bakar Alternatif Bersih (CAFV), Kelayakan Jangkauan Listrik, Harga, dll. Model machine learning mampu menangkap pola-pola tersembunyi yang sering luput dari sistem estimasi tradisional. Hasilnya, prediksi jarak tempuh menjadi lebih personal, adaptif, dan sesuai dengan kondisi nyata kendaraan maupun lingkungannya. Akurasi yang lebih tinggi ini diharapkan dapat mengurangi rasa cemas pengguna sekaligus meningkatkan kepercayaan diri mereka dalam menggunakan kendaraan listrik untuk aktivitas harian.

Berbagai penelitian telah dilakukan untuk memprediksi efisiensi kendaraan listrik menggunakan pendekatan *machine learning*. Penelitian oleh Saelan et al. (2026) mengembangkan model prediksi jarak tempuh kendaraan listrik menggunakan kombinasi Random Forest dan XGBoost, meskipun model mampu meningkatkan akurasi yang dihasilkan menunjukkan nilai akurasi R^2 sebesar 0,978, penelitian ini masih memiliki beberapa keterbatasan. Pertama, jumlah data yang digunakan relatif terbatas, yaitu hanya 478 data, sehingga berpotensi memengaruhi kemampuan generalisasi model. Kedua, model hanya berfokus pada satu pendekatan utama, yaitu kombinasi Random Forest dan XGBoost, tanpa melakukan perbandingan dengan algoritma lain, sehingga sulit untuk memastikan bahwa model tersebut merupakan yang paling optimal[7]. Selain itu, penelitian oleh Idris dan Pontoiyo (2025) pada penelitiannya menunjukkan bahwa metode Gradient Boosting mampu menghasilkan performa yang tinggi dengan nilai R^2 sebesar 0.963, masih terdapat beberapa keterbatasan. Pertama, penelitian hanya menggunakan dua metrik evaluasi, yaitu Mean Squared Error (MSE) dan R-squared (R^2), sehingga belum memberikan gambaran menyeluruh terhadap tingkat kesalahan model, seperti melalui Mean Absolute Error atau Root Mean Squared Error. Kedua, penelitian belum melakukan eksplorasi secara mendalam terhadap optimasi hyperparameter, yang diakui masih memerlukan perbaikan, sehingga performa model berpotensi belum optimal[8].

Sejumlah penelitian sebelumnya telah berupaya memprediksi jarak tempuh kendaraan listrik dengan memanfaatkan algoritma Machine Learning[9]. Walaupun pendekatan ini menunjukkan potensi dalam meningkatkan ketepatan estimasi, sebagian besar studi cenderung menggunakan satu algoritma secara langsung tanpa melalui tahapan seleksi fitur yang terstruktur. Konsekuensinya, model yang dihasilkan berpotensi memuat fitur yang kurang relevan atau bersifat redundan, sehingga menurunkan stabilitas, meningkatkan risiko overfitting, serta membatasi kemampuan untuk menafsirkan kontribusi masing-masing variabel input. Padahal, proses identifikasi fitur yang paling berpengaruh sangatlah krusial agar model prediktif yang dibangun tidak hanya akurat, tetapi juga mudah dipahami dan dapat dimanfaatkan secara praktis oleh pengguna maupun pengembang sistem. Dengan demikian, terdapat celah penelitian yang cukup jelas, yaitu minimnya pendekatan prediktif jarak tempuh kendaraan listrik yang mengintegrasikan pemilihan fitur secara sistematis dengan model prediksi yang kuat. Kesenjangan ini membuka peluang untuk mengembangkan kombinasi antara proses seleksi fitur yang terstruktur dan algoritma prediksi canggih, sehingga mampu menghasilkan estimasi jarak tempuh yang lebih stabil, akurat, serta adaptif terhadap kondisi penggunaan nyata.

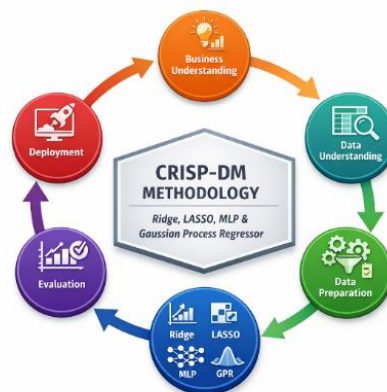
Meskipun penelitian-penelitian tersebut menunjukkan potensi penggunaan algoritma *machine learning*, sebagian besar hanya berfokus pada satu metode tanpa melakukan perbandingan secara komprehensif. Hal ini menyebabkan sulitnya menentukan model terbaik yang sesuai dengan karakteristik data. Selain itu, masih terbatas penelitian yang menggunakan pendekatan sistematis dalam proses data mining, seperti *Cross-Industry Standard Process for Data Mining* (CRISP-DM), untuk memastikan bahwa seluruh tahapan analisis dilakukan secara terstruktur. Berdasarkan hal tersebut, terdapat kesenjangan penelitian berupa kurangnya analisis komparatif yang terstruktur terhadap berbagai algoritma regresi dalam konteks prediksi efisiensi kendaraan listrik serta minimnya kajian mendalam terhadap problematika matematis dataset kendaraan listrik, khususnya terkait multikolinearitas, distribusi data yang tidak seimbang, dan trade-off antara akurasi serta interpretabilitas model. Penelitian dilakukan untuk membandingkan kinerja beberapa algoritma regresi Ridge Regression, LASSO, Multi-Layer Perceptron (MLP), dan Gaussian Process Regressor (GPR) untuk meningkatkan akurasi estimasi jarak tempuh kendaraan listrik. Ridge Regression biasa digunakan untuk mengatasi multikolinearitas dan menjaga stabilitas model dengan menekan koefisien variabel yang kurang relevan[10]. LASSO berperan dalam seleksi fitur dengan menyingkirkan variabel yang tidak signifikan, sehingga model lebih sederhana dan interpretable[11]. Sementara itu, MLP dipilih sebagai model prediksi utama karena kemampuannya menangkap pola non-linear yang kompleks dalam data kendaraan listrik, termasuk interaksi antar variabel yang sulit ditangkap oleh model linier[12]. Untuk melengkapi pendekatan ini, GPR digunakan sebagai model probabilistik yang tidak hanya menghasilkan prediksi akurat tetapi juga memberikan estimasi ketidakpastian, sehingga lebih adaptif terhadap

variasi kondisi penggunaan aktual[13]. Dengan menggabungkan kekuatan Ridge dan LASSO dalam pemilihan serta regularisasi fitur, lalu memanfaatkan kemampuan prediktif MLP dan GPR, model yang dihasilkan diharapkan mampu memberikan estimasi jarak tempuh kendaraan listrik yang lebih akurat, stabil, dan interpretable, baik dari sisi komputasional maupun praktis[14].

Pemilihan algoritma berdasar problematika matematis dataset kendaraan listrik, yaitu adanya multikolinearitas antar variabel teknis, distribusi data yang tidak seimbang, serta perbedaan skala antar fitur. Ridge dan LASSO dipilih karena mampu mengatasi multikolinearitas dan melakukan regularisasi, MLP dipertimbangkan untuk menangkap pola non-linear meskipun berisiko overfitting, sedangkan GPR relevan untuk memberikan estimasi probabilistik yang memperhitungkan ketidakpastian. Dengan demikian, analisis komparatif ini tidak hanya bersifat praktis, tetapi juga menjawab tantangan matematis yang melekat pada karakteristik data kendaraan listrik. Penelitian ini menggunakan pendekatan CRISP-DM yang meliputi tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Dengan pendekatan ini, proses analisis dilakukan secara sistematis dan terstruktur sehingga menghasilkan model yang lebih optimal. Tujuan dari penelitian ini adalah menganalisis kinerja beberapa algoritma regresi dalam prediksi efisiensi kendaraan listrik, membandingkan tingkat akurasi dan kesalahan dari masing-masing model, serta menentukan algoritma terbaik yang dapat digunakan dalam kasus prediksi efisiensi kendaraan listrik. Kontribusi utama dari penelitian ini bukan hanya pada analisis komparatif yang terstruktur menggunakan CRISP-DM, tetapi juga pada eksplorasi problematika matematis dataset kendaraan listrik dan bagaimana karakteristik tersebut memengaruhi performa algoritma. Hasil penelitian ini diharapkan dapat memberikan rekomendasi dalam pemilihan model yang tepat serta menjadi referensi bagi pengembangan sistem prediksi efisiensi kendaraan listrik di masa mendatang.

2. METODOLOGI PENELITIAN

Penelitian ini menerapkan metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*), sebuah kerangka kerja yang umum digunakan dalam analisis data dan pengembangan model prediktif berbasis *Machine Learning*[15]. Pendekatan ini dipilih karena sesuai untuk penelitian dengan data tabular, khususnya yang berfokus pada pembuatan model prediksi maupun klasifikasi berdasarkan atribut teknis. Gambar 1 menunjukkan tahapan metodologi penelitian yang dilakukan.



Gambar 1. Tahapan metodologi penelitian

2.1 Business Understanding

Pada tahap ini dilakukan beberapa kegiatan penting, seperti memahami kebutuhan serta tujuan penelitian dari perspektif bisnis, kemudian menerjemahkan pemahaman tersebut ke dalam bentuk perumusan masalah yang relevan dengan proses *data mining*. Permasalahan utama yang ingin dipecahkan adalah *range anxiety* atau kecemasan pengguna kendaraan listrik terhadap ketidakpastian estimasi jarak tempuh. Tujuan penelitian ini adalah membangun model prediksi jarak tempuh kendaraan listrik berdasarkan data populasi EV yang tersedia.

2.2 Data Understanding

Tahapan ini diawali dengan mengumpulkan data, mendeskripsikan data, serta mengevaluasi kualitas data. Dataset yang digunakan dalam penelitian ini berasal dari *open source kaggle* dengan judul *Electric Vehicle Population Data* yang berisi ribuan entri kendaraan listrik dengan berbagai atribut, seperti VIN (1-10), County, City, State, Postal Code, Model Year, Make, Model, Electric Vehicle Type, Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric Range, Base MSRP, Legislative District, DOL Vehicle ID, Vehicle Location, Electric Utility, 2020 Census Tract. Pada tahap ini dilakukan eksplorasi awal untuk memahami distribusi data, mengidentifikasi variabel numerik maupun kategorikal, serta mendeteksi adanya nilai yang hilang. Berikut pada Tabel 1 merupakan Deskripsi dataset.

Tabel 1. Deskripsi Dataset

Attributes	Description
------------	-------------

VIN (1–10)	Nomor identifikasi kendaraan (Vehicle Identification Number), digunakan untuk membedakan setiap unit kendaraan.
County	Nama wilayah administratif (county) tempat kendaraan terdaftar.
City	Kota tempat kendaraan berada atau terdaftar.
State	Negara bagian tempat kendaraan terdaftar.
Postal Code	Kode pos lokasi kendaraan.
Model Year	Tahun produksi kendaraan listrik.
Make	Merek kendaraan (misalnya Tesla, Nissan, Chevrolet).
Model	Nama model kendaraan listrik.
Electric Vehicle Type	Jenis kendaraan listrik (Battery Electric Vehicle/BEV atau Plug-in Hybrid Electric Vehicle/PHEV).
Clean Alternative Fuel Vehicle (CAFEV) Eligibility	Status kelayakan kendaraan untuk program bahan bakar alternatif bersih.
Electric Range	Jarak tempuh listrik maksimum (dalam km atau mil) dengan sekali pengisian penuh.
Base MSRP	Harga dasar kendaraan (Manufacturer's Suggested Retail Price).
Legislative District	Distrik legislatif tempat kendaraan terdaftar.
DOL Vehicle ID	Nomor identifikasi kendaraan dari Department of Licensing.
Vehicle Location	Lokasi kendaraan berdasarkan data registrasi.
Electric Utility	Nama perusahaan penyedia listrik yang melayani lokasi kendaraan.
2020 Census Tract	Kode wilayah sensus 2020 yang terkait dengan lokasi kendaraan.

2.3 Data Preparation

Pada tahap ini dibentuk dataset akhir yang berasal dari data mentah. Proses yang dilakukan mencakup pembersihan data (*data cleaning*) untuk menghilangkan inkonsistensi, pemilihan data (*data selection*) guna menentukan record dan atribut yang relevan, serta transformasi data (*data transformation*) agar data siap digunakan sebagai masukan dalam tahap pemodelan. Pembagian dataset dilakukan dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Praktik ini umum digunakan dalam *machine learning* pada dataset berukuran menengah, sehingga dapat menjaga kestabilan proses pelatihan sekaligus meningkatkan reliabilitas evaluasi model.

2.4 Modeling

Pada tahap ini, penelitian secara langsung melibatkan penerapan *Machine Learning* untuk menentukan teknik analisis data, perangkat bantu, serta algoritma yang digunakan. Empat model regresi dipilih untuk dibandingkan, yaitu Ridge Regression, LASSO Regression, Multi-Layer Perceptron (MLP), dan Gaussian Process Regressor (GPR). Setiap algoritma memiliki mekanisme berbeda dalam membangun model prediksi efisiensi kendaraan listrik:

a. Ridge Regression

Ridge Regression merupakan salah satu teknik regresi linier yang menggunakan regularisasi L2. Regularisasi ini bekerja dengan menambahkan penalti berupa kuadrat dari besarnya koefisien ke dalam fungsi loss, sehingga model tidak hanya berfokus pada meminimalkan error prediksi tetapi juga menjaga agar koefisien regresi tidak terlalu besar [16]. Dengan cara ini, Ridge Regression mampu mengurangi risiko overfitting, terutama pada dataset dengan jumlah variabel yang banyak atau ketika terdapat multikolinearitas (hubungan antar variabel independen yang sangat kuat). Dalam konteks penelitian ini, Ridge Regression digunakan untuk memprediksi efisiensi kendaraan listrik dengan mempertimbangkan variabel teknis seperti Range Efficiency, Model Year, Vehicle Age, dan Electric Vehicle Type. Keunggulan utama Ridge Regression adalah stabilitas model pada data berdimensi tinggi, sehingga hasil prediksi tetap konsisten meskipun terdapat variabel yang saling berkorelasi.

b. LASSO Regression

Least Absolute Shrinkage and Selection Operator (LASSO) merupakan metode regresi linier yang menggunakan regularisasi L1. Berbeda dengan Ridge Regression yang menekan koefisien melalui penalti kuadrat (L2), LASSO menambahkan penalti berupa nilai absolut dari koefisien ke dalam fungsi loss. Mekanisme ini tidak hanya mengurangi overfitting, tetapi juga berfungsi sebagai seleksi fitur otomatis, karena penalti L1 dapat mendorong sebagian koefisien menjadi nol [17]. Dengan demikian, variabel yang kurang relevan akan dieliminasi dari model, sehingga meningkatkan interpretabilitas dan menyederhanakan struktur model. Dalam konteks penelitian ini, LASSO Regression digunakan untuk memprediksi efisiensi kendaraan listrik dengan mempertahankan variabel dominan seperti Range Efficiency dan Model Year, sekaligus mengurangi kompleksitas model dengan mengeliminasi variabel administratif yang memiliki korelasi rendah. Keunggulan utama LASSO adalah kemampuannya menghasilkan model yang ringkas, akurat, dan mudah diinterpretasikan, sehingga sangat relevan untuk penelitian yang menekankan keterhubungan langsung antara variabel teknis dan efisiensi kendaraan [8].

c. Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP) merupakan salah satu bentuk jaringan saraf tiruan terdiri atas lapisan input, satu atau lebih lapisan tersembunyi (*hidden layers*), dan lapisan output. Setiap neuron dalam lapisan terhubung dengan neuron di lapisan berikutnya melalui bobot (*weights*) yang akan diperbarui selama proses pelatihan [18]. MLP menggunakan fungsi aktivasi non-linear seperti sigmoid, tanh, atau ReLU untuk memungkinkan jaringan menangkap hubungan

kompleks yang tidak dapat dimodelkan oleh regresi linier tradisional. Keunggulan utama MLP adalah kemampuannya dalam menangkap pola non-linear dan interaksi antar variabel yang kompleks[19]. Dalam penelitian ini, MLP menggunakan arsitektur dengan dua hidden layers berukuran 100 dan 50 neuron, fungsi aktivasi ReLU, serta optimizer Adam dengan maksimum iterasi 1000. Konfigurasi ini dipilih untuk menyeimbangkan kemampuan menangkap pola non-linear dengan risiko overfitting pada dataset berukuran ribuan entri.

d. Gaussian Process Regressor (GPR)

Gaussian Process Regressor (GPR) merupakan pendekatan non-parametrik dalam regresi yang memodelkan distribusi probabilistik atas fungsi yang ingin diprediksi. Berbeda dengan regresi linier klasik yang menetapkan bentuk fungsi tertentu, GPR mengasumsikan bahwa data berasal dari suatu proses Gaussian multivariat, sehingga setiap titik prediksi tidak hanya menghasilkan nilai estimasi, tetapi juga distribusi ketidakpastian[20]. Hal ini menjadikan GPR sangat berguna dalam konteks pengambilan keputusan berbasis risiko, karena model tidak hanya memberikan prediksi tunggal, tetapi juga interval kepercayaan yang menunjukkan tingkat kepastian hasil. Model GPR menggunakan kernel Radial Basis Function (RBF) dengan panjang skala (length scale) 1.0 yang dikalikan dengan Constant Kernel bernilai 1.0. Parameter regularisasi alpha ditetapkan sebesar $1e-2$ untuk mengurangi sensitivitas terhadap noise. Pemilihan kernel ini bertujuan untuk menangkap pola halus pada data kendaraan listrik sekaligus menjaga stabilitas perhitungan.

2.5 Evaluation

Model yang telah dilatih kemudian diuji menggunakan sejumlah metrik kuantitatif, yaitu Mean Absolute Error (MAE) untuk menilai rata-rata kesalahan absolut, Root Mean Squared Error (RMSE) untuk mengukur rata-rata kesalahan kuadrat yang lebih sensitif terhadap outlier, serta R-squared (R^2) yang menunjukkan proporsi variasi variabel target yang dapat dijelaskan oleh model. Selain analisis numerik, dilakukan pula evaluasi visual untuk memperkuat interpretasi hasil. Visualisasi mencakup scatter plot antara nilai aktual dan prediksi guna melihat kedekatan hasil estimasi, histogram residual untuk memahami distribusi galat, grafik kepentingan fitur untuk menilai kontribusi relatif tiap variabel, serta heatmap korelasi antar fitur numerik untuk mengamati hubungan linier antar variabel. Analisis visual ini berperan penting dalam menilai akurasi sekaligus interpretabilitas model secara menyeluruh. Selain pembagian data 80:20 untuk train-test, penelitian ini menerapkan prosedur k-fold cross-validation dengan $k=5$, menggunakan pengacakan (shuffle=True) dan random_state=42 untuk memastikan replikasi hasil. Evaluasi dilakukan dengan menghitung rata-rata RMSE dari setiap fold, sehingga hasil yang diperoleh lebih robust dibandingkan hanya menggunakan satu kali pembagian data.

2.6 Deployment

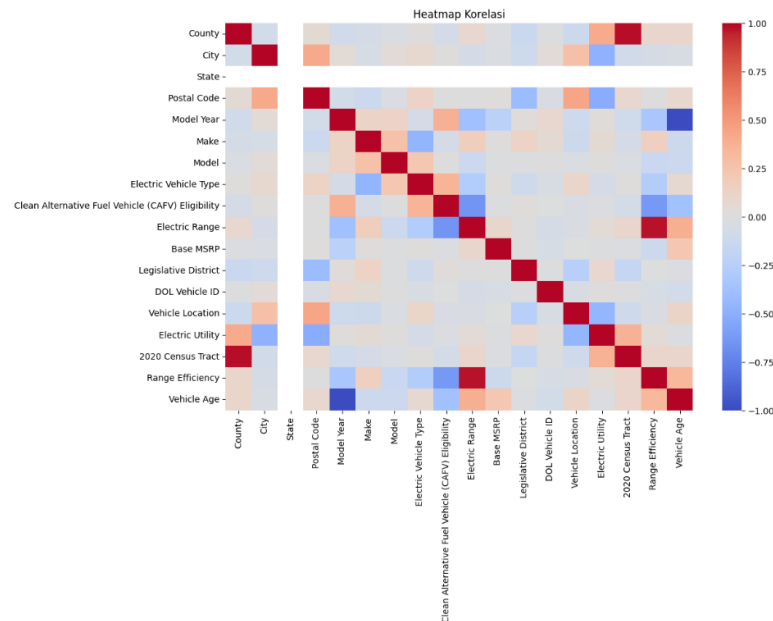
Model yang dihasilkan dalam penelitian ini dapat dianggap sebagai bukti konsep untuk sistem prediksi efisiensi kendaraan listrik berdasarkan data populasi dan atribut teknis yang tersedia. Dengan tingkat akurasi yang tinggi serta kemampuan generalisasi yang baik, model komparatif ini berpotensi diintegrasikan ke dalam sistem informasi kendaraan, aplikasi analitik efisiensi energi, maupun dashboard rekomendasi bagi konsumen dalam memilih kendaraan sesuai kebutuhan harian. Penelitian ini diharapkan dapat dijadikan dasar bagi pengembangan sistem pendukung keputusan bagi produsen kendaraan listrik dalam merancang produk yang lebih sesuai dengan tuntutan pasar. Tahap deployment berfokus pada dokumentasi model, visualisasi hasil, serta pemetaan arah implementasi di masa mendatang dalam skenario nyata, baik melalui sistem berbasis web maupun modul perangkat lunak analisis kendaraan. Dengan demikian, meskipun penelitian ini tidak secara langsung mengembangkan perangkat lunak atau platform aplikasi, pendekatan yang digunakan telah mempertimbangkan relevansi praktis dari model yang dibangun, sekaligus membuka peluang untuk adopsi lebih lanjut dan pengembangan sistem pada fase pasca penelitian. Dengan metodologi CRISP-DM memastikan penelitian berjalan sistematis, mulai dari perumusan masalah hingga validasi hasil, sehingga mendukung kredibilitas model yang dihasilkan serta memungkinkan replikasi dan pengembangan lebih lanjut[21].

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data

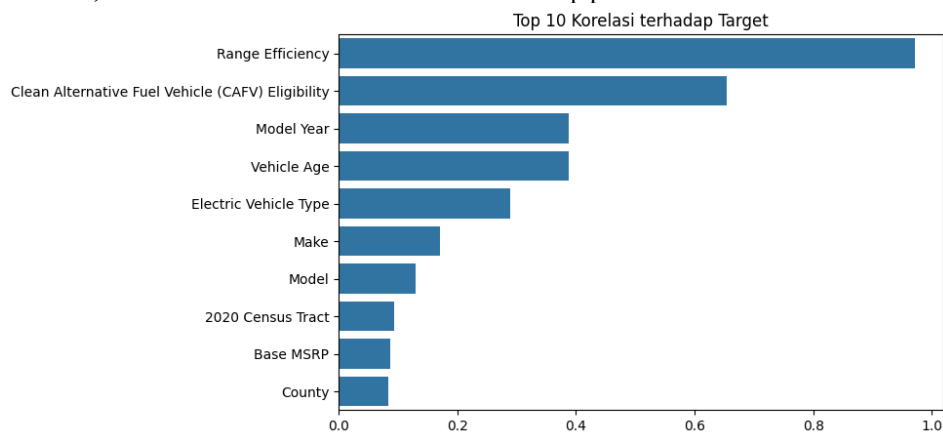
Data dalam penelitian diambil dari dataset Electric Vehicle Population yang mewakili spesifikasi kendaraan listrik dengan beragam atribut teknis dan demografis. Setiap entri mencakup informasi seperti VIN (1-10), County, City, State, Postal Code, Model Year, Make, Model, Electric Vehicle Type, Clean Alternative Fuel Vehicle (CAFEV) Eligibility, Electric Range, Base MSRP, Legislative District, DOL Vehicle ID, Vehicle Location, Electric Utility, 2020 Census Tract. Dengan pra-pemrosesan dan imputasi nilai yang hilang, data kategorikal diubah menjadi representasi numerik menggunakan teknik Label Encoding maupun One-Hot Encoding, sehingga dapat diproses oleh algoritma regresi. Transformasi ini memungkinkan integrasi variabel kategorikal ke dalam model prediktif tanpa menambah kompleksitas dimensi secara berlebihan. Hasil evaluasi menunjukkan bahwa proses encoding tidak menurunkan kinerja model, yang tercermin dari nilai R^2 yang tinggi pada sebagian besar algoritma. Hal ini mengindikasikan bahwa model mampu mempelajari pola hubungan antara variabel input dan target efisiensi kendaraan listrik secara konsisten. Selanjutnya, dilakukan seleksi fitur berbasis korelasi dan uji multikolinearitas untuk mengidentifikasi variabel yang paling relevan terhadap efisiensi kendaraan. Fitur-fitur dengan tingkat kepentingan tertinggi digunakan dalam pelatihan model Ridge, LASSO, MLP, dan Gaussian Process Regressor, sehingga kompleksitas model dapat dikurangi dan fokus diarahkan pada variabel yang paling berpengaruh terhadap efisiensi energi kendaraan listrik. Proses pemilihan fitur dalam penelitian ini

dilakukan untuk memastikan bahwa model yang dibangun berfokus pada variabel yang paling relevan terhadap efisiensi kendaraan listrik. Seleksi fitur dilakukan dengan pendekatan berbasis analisis korelasi, regularisasi, dan uji multikolinearitas, sehingga variabel yang memiliki kontribusi terbesar terhadap target dapat diidentifikasi.



Gambar 2. Heatmap Matriks Korelasi

Pada Gambar 2 merupakan heatmap yang menampilkan hubungan linier (korelasi) antar variabel dalam dataset. Sumbu horizontal dan vertikal. Dengan demikian, heatmap ini berfungsi sebagai alat eksplorasi awal untuk memahami hubungan antar variabel, membantu dalam seleksi fitur sebelum tahap pemodelan.



Gambar 3. Top 10 Korelasi terhadap target

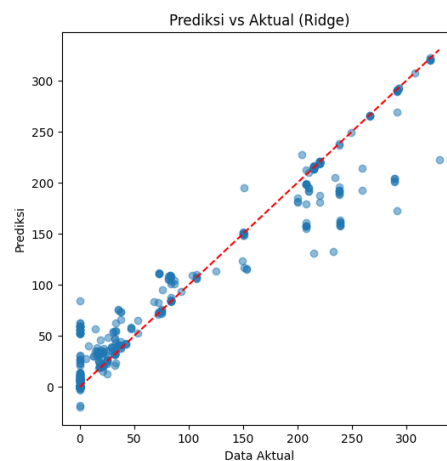
Pada Gambar 3 terdapat grafik yang membantu mengidentifikasi variabel yang paling relevan untuk dimasukkan ke dalam model prediksi. Hasil visualisasi menunjukkan bahwa Range Efficiency memiliki korelasi paling tinggi, mendekati nilai 1.0, sehingga menjadi variabel dominan dalam memengaruhi efisiensi kendaraan listrik. Variabel lain seperti CAFV Eligibility, Model Year, dan Vehicle Age juga memiliki korelasi kuat, menegaskan bahwa kelayakan program bahan bakar alternatif, tahun produksi, serta usia kendaraan berperan penting dalam menentukan efisiensi. Selain itu, variabel teknis seperti Electric Vehicle Type, Make, dan Model menunjukkan hubungan sedang hingga kuat, mencerminkan peran jenis kendaraan, produsen, dan desain model terhadap efisiensi. Sementara itu, faktor administratif seperti 2020 Census Tract, Base MSRP, dan County memiliki korelasi lebih rendah, sehingga kontribusinya terhadap efisiensi relatif kecil. Secara keseluruhan, diagram ini menegaskan bahwa atribut teknis kendaraan lebih dominan dibanding faktor administratif dalam memengaruhi efisiensi kendaraan listrik. Temuan ini menjadi dasar penting dalam proses seleksi fitur untuk pemodelan prediktif, sehingga algoritma regresi dapat difokuskan pada variabel yang paling relevan terhadap performa kendaraan.

Tabel 2. Top 10 Korelasi Variabel

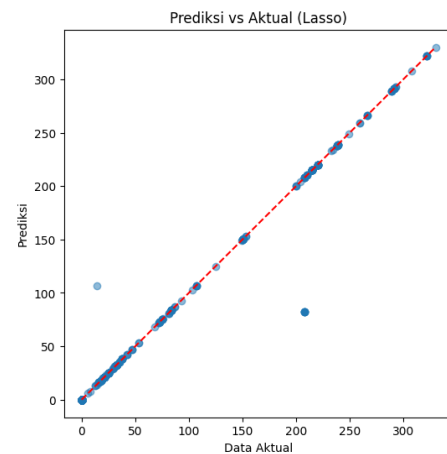
No.	Variabel	Nilai Korelasi ($\approx$)
1	Range Efficiency	~ 0.95

2	CAFV Eligibility	~0.80
3	Model Year	~0.75
4	Vehicle Age	~0.70
5	Electric Vehicle Type	~0.65
6	Make	~0.60
7	Model	~0.55
8	2020 Census Tract	~0.45
9	Base MSRP	~0.40
10	County	~0.35

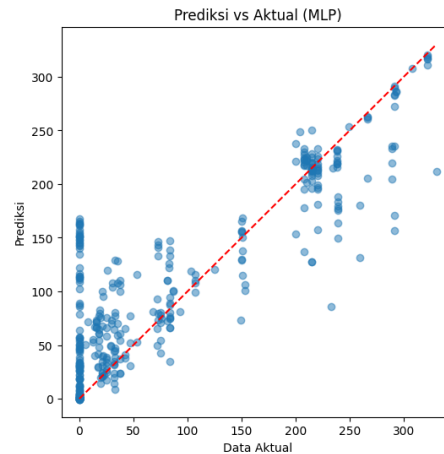
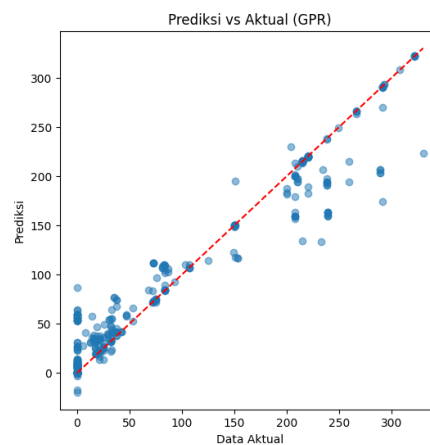
Nilai korelasi variabel lebih detail dapat dilihat pada Tabel 2 analisis ini menegaskan bahwa atribut teknis kendaraan lebih dominan dibandingkan faktor administratif dalam memengaruhi efisiensi kendaraan listrik. Oleh karena itu, variabel teknis menjadi prioritas dalam proses seleksi fitur untuk pemodelan prediktif. Temuan ini memberikan dasar yang kuat bagi penggunaan algoritma regresi seperti Ridge, LASSO, MLP, dan Gaussian Process Regressor, sehingga model dapat difokuskan pada variabel yang paling relevan dan menghasilkan prediksi efisiensi yang lebih akurat serta stabil.



Gambar 4. Visualiasi Prediksi vs Data Aktual Ridge



Gambar 5. Visualiasi Prediksi vs Data Aktual LASSO

**Gambar 6.** Visualiasi Prediksi vs Data Aktual MLP**Gambar 7.** Visualiasi Prediksi vs Data Aktual Gaussian Process Regressor

Berdasarkan Gambar 4, 5, 6, dan 7 menunjukkan Visualiasi Prediksi vs Data Aktual Ridge, Lasso, MLP, dan GPR. Keempat gambar menunjukkan bahwa model Ridge, LASSO, MLP, dan GPR mampu menghasilkan prediksi yang cukup dekat dengan data aktual. Ridge dan LASSO lebih stabil dengan sebaran titik rapat, MLP unggul dalam menangkap pola non-linear, sedangkan GPR memberikan prediksi probabilistik dengan estimasi ketidakpastian. Ridge Regression menggunakan regularisasi L2 untuk menjaga kestabilan model dan mengatasi multikolinearitas, sehingga tetap konsisten meski terdapat variabel yang saling berkorelasi. LASSO Regression dengan regularisasi L1 berfungsi sebagai seleksi fitur, mempertahankan variabel dominan seperti Range Efficiency dan Model Year, sehingga model lebih sederhana, akurat, dan mudah diinterpretasikan. MLP mampu memodelkan hubungan kompleks antar variabel teknis seperti harga dasar kendaraan dan usia kendaraan, namun membutuhkan komputasi lebih tinggi dan menghasilkan error lebih besar dibanding Ridge dan LASSO. GPR dengan kernel RBF memberikan prediksi cukup akurat sekaligus menyajikan tingkat kepercayaan, sehingga lebih informatif untuk pengambilan keputusan berbasis risiko.

3.2 Hasil

Setiap algoritma diuji dengan konfigurasi awal sebagai baseline, lalu dilakukan tuning parameter dan validasi silang untuk meningkatkan akurasi. Pendekatan ini memungkinkan evaluasi menyeluruh atas keunggulan dan keterbatasan masing-masing algoritma dalam memprediksi efisiensi kendaraan listrik.

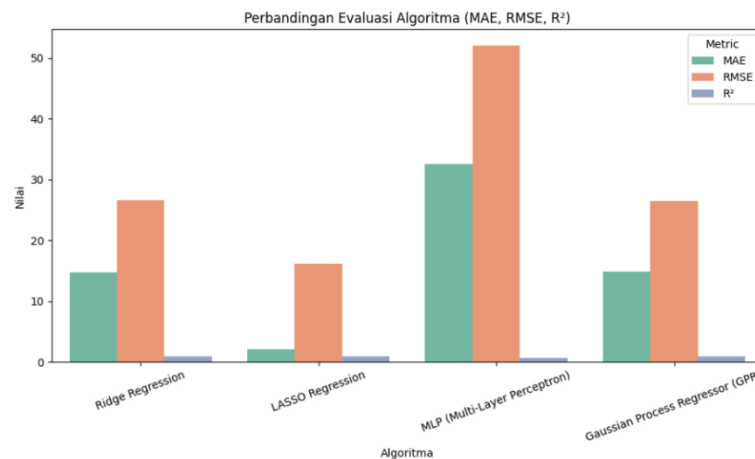
Tabel 3. Perbandingan Evaluasi Model

Algoritma	MAE	RMSE	R ²
Ridge Regression	14.75	26.60	0.929
LASSO Regression	2.13	16.10	0.974
MLP (Multi-Layer Perceptron)	32.57	52.09	0.729
Gaussian Process Regressor (GPR)	14.89	26.52	0.930

Tabel 3 ditampilkan perbandingan hasil evaluasi model pada penelitian ini. Proses pengoptimalan hyperparameter pada keempat algoritma menghasilkan peningkatan performa model yang ditunjukkan melalui penurunan nilai Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE), serta peningkatan nilai R² yang merepresentasikan kesesuaian model dengan data. Menunjukkan bahwa setiap algoritma memiliki keunggulan dan

keterbatasan masing-masing. LASSO Regression memberikan nilai R^2 tertinggi (0.974) dan MAE paling rendah (2.13), menegaskan kemampuannya dalam melakukan seleksi fitur dan menangkap pola umum secara efektif. Namun, disparitas signifikan dengan RMSE sebesar 16.10 mengindikasikan adanya varian error besar atau outlier yang tidak sepenuhnya ditangani, sehingga interpretasi hasil perlu dilakukan dengan hati-hati. Ridge Regression dan Gaussian Process Regressor (GPR) menunjukkan performa yang relatif stabil dengan R^2 sekitar 0.93 dan nilai MAE serta RMSE yang seimbang. Ridge unggul dalam mengatasi multikolinearitas, sementara GPR memberikan tambahan informasi berupa estimasi ketidakpastian yang bermanfaat dalam konteks pengambilan keputusan berbasis risiko. Sebaliknya, Multi-Layer Perceptron (MLP) mampu menangkap pola non-linear antar variabel, tetapi menghasilkan error yang lebih tinggi (MAE 32.57, RMSE 52.09, R^2 0.729). Hal ini menunjukkan bahwa MLP kurang optimal untuk dataset berukuran ribuan entri, karena cenderung sensitif terhadap kompleksitas data dan berisiko overfitting.

Secara keseluruhan, LASSO dapat dianggap unggul dari sisi akurasi rata-rata dan interpretabilitas, Ridge dan GPR lebih konsisten serta stabil dalam menghadapi distribusi data yang kompleks, sedangkan MLP relevan untuk eksplorasi pola non-linear namun kurang efisien pada dataset ini.



Gambar 8. Visualisasi Hasil Evaluasi Model

Pada Gambar 8 menunjukkan visualisasi hasil evaluasi model bahwa pemilihan algoritma harus mempertimbangkan tujuan penelitian. LASSO Regression lebih unggul untuk akurasi dan interpretabilitas, Ridge Regression dan GPR memberikan kestabilan serta estimasi yang konsisten, sedangkan MLP relevan untuk memodelkan hubungan kompleks meskipun dengan biaya komputasi lebih tinggi. Dengan demikian, hasil komparatif ini memperkuat kesimpulan bahwa optimisasi hyperparameter berkontribusi pada prediksi efisiensi kendaraan listrik yang lebih akurat dan kuat, sekaligus memberikan dasar bagi pengembangan sistem prediktif berbasis data teknis kendaraan. Hasil penelitian menunjukkan bahwa fitur teknis kendaraan listrik memiliki hubungan yang kuat dengan efisiensi dan jarak tempuh aktual kendaraan. Visualisasi hasil prediksi memperlihatkan bahwa model yang dibangun memiliki kinerja yang stabil dan mampu memberikan estimasi jarak tempuh dengan tingkat akurasi tinggi. Temuan ini memperkuat kesimpulan bahwa pendekatan berbasis pemilihan fitur dan pemodelan regresi terkontrol efektif dalam konteks data spesifikasi teknis kendaraan listrik.

Secara khusus, variabel seperti Model Year, Vehicle Age, Range Efficiency, dan Electric Vehicle Type terbukti menjadi penentu utama dalam estimasi efisiensi kendaraan. Hal ini sejalan dengan ekspektasi logis bahwa kendaraan dengan teknologi baterai lebih baru dan efisiensi energi lebih tinggi akan memiliki jarak tempuh yang lebih panjang. Selain itu, atribut seperti Make dan Model juga memberikan kontribusi signifikan terhadap performa prediksi, mencerminkan perbedaan desain dan teknologi antar produsen. Fitur-fitur tersebut memberikan wawasan penting bahwa karakteristik teknis dan efisiensi energi berperan besar dalam kinerja kendaraan listrik. Hasil komparatif menunjukkan bahwa LASSO Regression memberikan akurasi tertinggi dengan nilai R^2 sebesar 0.974, diikuti oleh Ridge Regression dan Gaussian Process Regressor yang menunjukkan kestabilan model dengan R^2 sekitar 0.93. Sementara itu, MLP mampu menangkap pola non-linear namun menghasilkan error yang lebih tinggi.

Dengan demikian, kombinasi antara analisis korelasi fitur dan pemodelan regresi komparatif terbukti efektif untuk memahami faktor-faktor yang memengaruhi efisiensi kendaraan listrik. Pendekatan ini dapat menjadi dasar pengembangan sistem prediksi jarak tempuh berbasis data teknis kendaraan, baik untuk aplikasi industri otomotif maupun sistem rekomendasi bagi pengguna kendaraan listrik. Penelitian ini membuka peluang untuk mengintegrasikan data teknis kendaraan listrik dengan data pola penggunaan di dunia nyata, seperti informasi GPS, suhu lingkungan, serta kebiasaan pengisian daya. Integrasi tersebut dapat menghasilkan model hibrida prediktif yang lebih komprehensif dan adaptif terhadap kondisi aktual. Selain itu, pengembangan sistem rekomendasi berbasis spesifikasi teknis dan efisiensi kendaraan menjadi arah penelitian yang menarik di masa mendatang. Pendekatan ini tidak hanya relevan bagi pengguna kendaraan listrik dalam memahami performa harian, tetapi juga bermanfaat bagi produsen dan pengambil kebijakan dalam merancang strategi pengembangan kendaraan listrik yang lebih efisien. Dengan demikian, penelitian ini menegaskan bahwa kombinasi analisis komparatif algoritma Ridge, LASSO, MLP, dan Gaussian Process Regressor dapat menjadi

fondasi penting bagi pengembangan sistem prediksi efisiensi kendaraan listrik, sekaligus membuka ruang bagi eksplorasi lebih lanjut di bidang kecerdasan buatan dan mobilitas berkelanjutan.

3.3. Pembahasan

Hasil penelitian ini menunjukkan bahwa LASSO Regression memberikan akurasi tertinggi ($R^2 = 0.97$) dengan kemampuan seleksi fitur yang kuat. Temuan ini sejalan dengan studi oleh Saelan et al. (2026) yang melaporkan efektivitas kombinasi Random Forest feature selection dan XGBoost dalam meningkatkan akurasi prediksi jarak tempuh kendaraan listrik ($R^2 = 0.978$, MAE = 10.55) meskipun penelitian tersebut menggunakan dataset yang lebih kecil (478 entri) dibandingkan dengan Electric Vehicle Population Data yang digunakan dalam studi ini[7]. Ridge Regression dan Gaussian Process Regressor dalam penelitian ini juga konsisten dengan hasil Putri et al. (2025) yang menekankan kestabilan model regresi terkontrol dalam konteks prediksi biaya asuransi kesehatan ($R^2 = 0.857$), menunjukkan bahwa regularisasi L2 mampu menjaga stabilitas meskipun data memiliki multikolinearitas[22]. Sebaliknya, hasil MLP pada penelitian ini ($R^2 = 0.72$) berbeda dengan temuan Ahmadi et al. (2020) yang menemukan performa tinggi ANN dalam sistem kontrol diferensial kendaraan listrik dengan dataset terbatas; pada dataset besar dengan ribuan entri, MLP justru menghasilkan error lebih tinggi, menegaskan sensitivitas algoritma neural network terhadap kompleksitas data[23].

Dengan demikian, penelitian ini tidak hanya memperkuat temuan terdahulu mengenai efektivitas regresi terkontrol, tetapi juga memberikan kontribusi baru melalui penggunaan dataset Electric Vehicle Population Data yang lebih komprehensif serta integrasi analisis korelasi fitur. Hal ini menegaskan bahwa pemilihan algoritma harus mempertimbangkan ukuran dan kompleksitas dataset, serta tujuan penelitian, sehingga hasil prediksi efisiensi kendaraan listrik dapat lebih akurat dan stabil.

4. KESIMPULAN

Penelitian ini menegaskan bahwa analisis komparatif terhadap algoritma Ridge Regression, LASSO Regression, Multi-Layer Perceptron (MLP), dan Gaussian Process Regressor (GPR) dalam memprediksi efisiensi kendaraan listrik memberikan pemahaman yang lebih mendalam mengenai hubungan antara variabel teknis dan performa model. LASSO Regression menunjukkan akurasi tertinggi, namun disparitas antara MAE dan RMSE mengindikasikan sensitivitas terhadap outlier, sehingga interpretasi hasil dengan berkala. Ridge Regression dan GPR terbukti lebih stabil, dengan error yang seimbang dan R^2 mendekati 0.93, menjadikannya pilihan yang lebih konsisten untuk data dengan distribusi kompleks. Sementara itu, MLP mampu menangkap pola non-linear, tetapi gagal memodelkan variabel teknis dominan seperti Range Efficiency dan Vehicle Age secara efektif, sehingga menghasilkan error lebih tinggi. Dari perspektif otomotif dan energi, temuan ini menegaskan bahwa efisiensi kendaraan listrik sangat dipengaruhi oleh faktor teknis seperti tahun produksi, usia kendaraan, dan kapasitas baterai. Model linear dengan mekanisme regularisasi Ridge dan LASSO lebih mampu menangkap pengaruh langsung variabel-variabel tersebut dibandingkan pendekatan berbasis jaringan saraf yang cenderung terjebak pada kompleksitas data. Hal ini memberikan insight penting bagi industri otomotif: pengembangan sistem prediksi efisiensi sebaiknya berfokus pada integrasi variabel teknis utama dengan model yang stabil, sehingga hasil prediksi dapat digunakan untuk strategi desain kendaraan dan kebijakan energi yang lebih akurat. Untuk penelitian mendatang, peta jalan yang lebih konkret mencakup: Integrasi data teknis dengan data dunia nyata (GPS, suhu lingkungan, kebiasaan pengisian daya) untuk membangun model hibrida yang lebih adaptif; Eksplorasi algoritma ensemble (misalnya kombinasi Ridge-LASSO atau LASSO-GPR) guna meningkatkan akurasi sekaligus ketahanan terhadap outlier; Analisis residual dan robust regression untuk memahami distribusi error serta memperkuat validitas model dalam konteks operasional kendaraan Listrik; atau Pengembangan sistem rekomendasi berbasis efisiensi energi yang dapat digunakan oleh konsumen maupun produsen untuk pengambilan keputusan strategis. Dengan demikian, penelitian ini tidak hanya membandingkan performa algoritma, tetapi juga memberikan wawasan praktis bagi industri otomotif dan energi mengenai faktor teknis yang paling berpengaruh terhadap efisiensi kendaraan listrik, sekaligus membuka arah penelitian lanjutan yang lebih terarah dan aplikatif.

REFERENCES

- [1] A. D. Irianto and K. U. Wibowo, "Pengembangan Inovasi Transit Oriented Development Dengan Electric Vehicle Sebagai Smart Mobility Pengurangan Emisi Dan Pembangunan Berkelanjutan Di Ibu Kota Nusantara," *Integr. Perspect. Soc. Sci. J.*, vol. 2, no. 1, pp. 1495–1508, 2025, [Online]. Available: <https://ipssj.com/index.php/ojs/article/view/199>
- [2] S. A. Ferlita, Sudarti, and Yushardi, "Analisis Efisiensi Kendaraan Listrik Sebagai Salah Satu Transportasi Ramah Lingkungan Pengurang Emisi Karbon," *Opt. J. Pendidik. Fis.*, vol. 7, no. 2, pp. 356–365, 2023, doi: <https://doi.org/10.37478/optika.v7i2.3282>.
- [3] P. Chakraborty et al., "Addressing the range anxiety of battery electric vehicles with charging en route," *Sci. Rep.*, pp. 1–15, 2022, doi: [10.1038/s41598-022-08942-2](https://doi.org/10.1038/s41598-022-08942-2).
- [4] B. Powell and C. Johnson, "Impact of Electric Vehicle Charging Station Reliability, Resilience, and Location on Electric Vehicle Adoption," *Natl. Renew. Energy Lab.*, no. August, 2024, doi: <https://doi.org/10.2172/2432350>.
- [5] S. Mondal, H. N. Ramesh, X. Ma, S. Cheng, T. Angeles, and S. Sun, "Reducing Electric Vehicle Range Anxiety with Trip-Specific Charge Usage Predictions," *ChemRxiv (pronounced 'chem-archive')*, no. March, pp. 0–28, 2025, doi: <https://doi.org/10.26434/chemrxiv-2024-ztnx9-v2>.
- [6] V. T. D. Sang, Q. H. Duong, L. Zhou, and C. F. A. Arranz, "Electric Vehicle Battery Technologies and Capacity Prediction: A Comprehensive Literature Review of Trends and Influencing Factors," *MDPI J.*, 2024, doi: <https://doi.org/10.3390/batteries10120451>.
- [7] M. R. R. Saelan, R. L. Hasanah, and S. Fauziah, "Optimalisasi Prediksi Jarak Tempuh Kendaraan Listrik Menggunakan Xgboost Dan Feature Selection Random Forest," *INTI NUSA MANDIRI*, vol. 20, no. 2, pp. 213–221, 2026, doi: <https://doi.org/10.33480/zahvvq98>.



- [8] N. O. Idris and F. Pontoioy, "Evaluasi Model Machine Learning untuk Prediksi Harga Mobil dengan Perbandingan Ensemble dan Regresi Linear," *J. Ilmu Komput. dan Sist. Inf.*, vol. 4, no. 1, pp. 129–143, 2025, doi: <https://doi.org/10.70340/jirsi.v4i1.181>.
- [9] I. Hussain, K. B. Ching, C. Uttrapphan, K. G. Tay, and A. Noor, "Evaluating machine learning algorithms for energy consumption prediction in electric vehicles: A comparative study," *Sci. Rep.*, pp. 1–20, 2025, doi: <https://doi.org/10.1038/s41598-025-94946-7>.
- [10] V. Esen, B. Coban, B. Y. K. T. K. Karaca, T. Dindar, and A. S. Sarkin, "An interpretable statistical approach to photovoltaic power forecasting using factor analysis and ridge regression," *Sci. Reports*, pp. 1–17, 2025, doi: <https://doi.org/10.1038/s41598-025-22838-x>.
- [11] M. A. Khan, K. Mahboob, U. Yousuf, M. Ramzan, M. T. Shaikh, and S. Akber, "Investigating the Role of LASSO in Feature Selection for Educational Data Mining (EDM) Applications," *VFAST Trans. Softw. Eng.*, vol. 13, no. 2, pp. 56–67, 2025, doi: <https://doi.org/10.21015/vtse.v13i2.2111>.
- [12] D. P. Mishra, P. Kumar, P. Rai, A. Kumar, and S. R. Salkuti, "Exploratory data analysis for electric vehicle driving range prediction : insights and evaluation," *Int. J. Appl. Power Eng.*, vol. 13, no. 2, pp. 474–482, 2024, doi: [10.11591/ijape.v13i2.pp474-482](https://doi.org/10.11591/ijape.v13i2.pp474-482).
- [13] V. L. Deringer, A. P. Bart, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, "Gaussian Process Regression for Materials and Molecules," *Chem. Rev.*, 2021, doi: [10.1021/acs.chemrev.1c00022](https://doi.org/10.1021/acs.chemrev.1c00022).
- [14] D. Kim, H. G. Shim, and J. S. Eo, "A Machine Learning Method for EV Range Prediction with Updates on Route Information and Traffic Conditions," *IAAI Tech. Track Emerg. Appl.*, 2022, doi: <https://doi.org/10.1609/aaai.v36i11.21525>.
- [15] R. Winurputra and D. E. Ratnawati, "Peramalan Penjualan Produk Menggunakan Extreme Gradient Boosting (Xgboost) Dan Kerangka Kerja Crisp-Dm Untuk Pengoptimalan Manajemen Persediaan (Studi Kasus: Ub Mart)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 2, pp. 417–428, 2025, doi: [10.25126/jtiik.2025129451](https://doi.org/10.25126/jtiik.2025129451).
- [16] A. F. Marwa, S. A. Setiyawan, Y. T. N. Cahyani, and H. D. Cahyono, "Optimization Of Stock Price Prediction With Ridge Regression And Hyperparameter Selections," *J. Tek. Inform.*, vol. 6, no. 1, pp. 141–148, 2025, doi: <https://doi.org/10.52436/1.jutif.2025.6.1.2384>.
- [17] A. Z. Panjaitan, A. Badawi, R. R. Rambe, and R. Audina, "Perbandingan Metode Kuadrat Terkecil , LASSO , Ridge , dan Elastic Net dalam Memprediksi Harga Mobil," *Reliab. J. Ind. Stat.*, vol. 01, no. 2, pp. 45–58, 2025, [Online]. Available: <https://journals.arcstics.id/reliability/article/view/25>
- [18] S. H. Gulo and A. H. Lubis, "Penerapan Multi-Layer Perceptron untuk Mengklasifikasi Penduduk Kurang Mampu," *J. Comput. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 51–59, 2024, doi: <https://doi.org/10.47065/explorer.v4i2.1146>.
- [19] D. Arifuddin, Kusni, and Kusnawi, "Perbandingan Performansi Algoritma Multiple Linear Regression dan Multi Layer Perceptron Neural Network dalam Memprediksi Penjualan Obat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. April, pp. 722–737, 2025, doi: <https://doi.org/10.57152/malcom.v5i2.1952>.
- [20] R. P. H. Siagian, A. D. Rahadjoe, and M. M. Al Haromainy, "Optimasi Hyperparameter Untuk Prediksi Nilai Tukar Mata Uang Dollar AS (USD) Dengan Rupiah (IDR) Menggunakan Gaussian Process Regression (GPR) Dengan Rational Quadratic dan Matern Kernel," *JiIP (Jurnal Ilm. Ilmu Pendidikan)*, vol. 8, no. 11, pp. 13033–13042, 2025, doi: <https://doi.org/10.54371/jiip.v8i11.9914>.
- [21] S. S. Abdul Muis, Havid Syafwan, Afif Zuhri Arfianto, Mutiara S. Simanjuntak, Albert Riyandi , Ahmad Budi Trisnawan, William Ramdhan, Helza Triana, Muhammad Hadi Saputra, Wiwin Handoko, Ahmad Naswin, *DATA MINING: Konsep, Metode, dan Aplikasi*. JAMBI: FAASLIB SERAMBI MEDIA, 2025. [Online]. Available: https://books.google.co.id/books?hl=id&lr=&id=j-WbEQAAQBAJ&oi=fnd&pg=PA76&dq=Metodologi+CRISP-DM+memastikan+penelitian+berjalan+sistematis,+mulai+dari+perumusan+masalah+hingga+validasi+hasil,+sehingga+mendukung+kredibilitas+model+yang+dihasilkan+serta+memungkinkan+replikasi+dan+pengembangan+lebih+lanjut&ots=qeP5V0rwh5&sig=KadqwG16auhBwRRKpCg8QnORAA&redir_esc=y#v=onepage&q&f=false
- [22] S. R. Putri, F. Supriadi, and D. Setiadi, "Perbandingan Model Regresi Nonlinear Polynomial , Ridge , dan Lasso untuk Prediksi Biaya Asuransi Kesehatan Berdasarkan Comparison of Nonlinear Regression Models — Polynomial , Ridge , and Lasso — for Predicting Health Insurance Costs Based on the CRISP-D," *TelKa (Teknologi Inf. dan Komunikasi)*, vol. 15, no. 0261, pp. 27–33, 2026, doi: <https://doi.org/10.36342/3kxrvj44>.
- [23] S. Ahmadi, K. Anam, and Widjonarko, "Peningkatan Efisiensi Energi pada Kendaraan Listrik dengan Elektronik Diferensial Berbasis ANN (Artificial Neural Network)," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 8, no. 3, pp. 642–656, 2020, doi: <http://dx.doi.org/10.26760/elkomika.v8i3.642>.