

OPTIMALISASI PREDIKSI JARAK TEMPUH KENDARAAN LISTRIK MENGUNAKAN XGBOOST DAN *FEATURE SELECTION* RANDOM FOREST

M. Rangga Ramadhan Saelan¹; Riyan Latifahul Hasanah^{2*}; Siti Fauziah³

Sains Data¹

Sistem Informasi^{2,3}

Universitas Nusa Mandiri, DKI Jakarta, Indonesia^{1,2,3}

www.nusamandiri.ac.id^{1,2,3}

rangga.mgg@nusamandiri.ac.id¹, riyan.rlt@nusamandiri.ac.id^{2*}, siti.suz@nusamandiri.ac.id³

(*) Corresponding Author



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

Abstract— *Electric vehicle (EV) adoption is increasing with increasing awareness of clean energy and environmental concerns. However, range anxiety, the uncertainty in estimating driving range, remains a major barrier. This study aims to develop a predictive model for EV range based on technical specifications to provide more accurate estimates and alleviate user concerns. A Machine Learning approach is applied, using Random Forest for feature selection and XGBoost as the primary prediction algorithm. The dataset consists of 478 EV records with 22 attributes, including battery_capacity, efficiency, dimensions, and speed. Key features affecting range prediction include battery_capacity_kWh, efficiency_wh_per_km, and height_mm. The XGBoost model demonstrates strong predictive performance with an R^2 of 0.978, an MAE of 10.555, and an RMSE of 15.180. These results suggest that combining Random Forest and XGBoost offers a promising solution to improve the accuracy of EV range estimation, potentially reducing range anxiety and supporting wider EV adoption.*

Keywords: *Electric Vehicle (EV), Machine Learning, Prediction, Random Forest, XGBoost*

Abstrak—Penggunaan kendaraan listrik (*Electric Vehicle*/EV) meningkat seiring dengan meningkatnya kesadaran akan energi bersih dan masalah lingkungan. Namun, kecemasan akan jangkauan, ketidakpastian dalam memperkirakan jangkauan berkendara, tetap menjadi hambatan utama. Studi ini bertujuan untuk mengembangkan model prediksi untuk jangkauan EV berdasarkan spesifikasi teknis untuk memberikan perkiraan yang lebih akurat dan mengurangi kekhawatiran pengguna. Pendekatan *Machine Learning* diterapkan, menggunakan *Random Forest* untuk pemilihan fitur dan *XGBoost* sebagai algoritma prediksi utama. Kumpulan data terdiri dari 478 *record* EV dengan 22 atribut, termasuk kapasitas baterai, efisiensi, dimensi, dan kecepatan. Fitur utama yang memengaruhi prediksi jangkauan meliputi *battery_capacity_kWh*, *efficiency_wh_per_km*, dan *height_mm*. Model *XGBoost* menunjukkan kinerja prediktif yang kuat dengan R^2 0,978, MAE 10,555, dan RMSE 15,180. Hasil ini menunjukkan bahwa menggabungkan *Random Forest* dan *XGBoost* menawarkan solusi yang menjanjikan untuk meningkatkan akurasi estimasi jangkauan EV, yang berpotensi mengurangi kecemasan akan jangkauan dan mendukung adopsi EV yang lebih luas.

Kata kunci: *Electric Vehicle (EV), Machine Learning, Prediksi, Random Forest, XGBoost*

PENDAHULUAN

Kendaraan listrik (*Electric Vehicle*/EV) merupakan salah satu alternatif transportasi masa depan yang ramah lingkungan dan berkelanjutan. Keunggulan utamanya terletak pada emisi karbon yang jauh lebih rendah dibandingkan kendaraan berbahan bakar fosil (Al-Ghaili et al., 2022; Sudjoko et al., 2021; Zhang et al., 2023). sehingga dapat

mengurangi dampak negatif terhadap lingkungan, terutama dalam hal polusi udara dan kontribusi terhadap pemanasan global. Teknologi EV juga mendukung transisi menuju sistem energi yang lebih bersih dengan memanfaatkan sumber energi terbarukan sebagai pengganti bahan bakar minyak (Sinha et al., 2023). Seiring meningkatnya kesadaran masyarakat terhadap isu perubahan iklim, kerusakan lingkungan, dan terbatasnya

sumber daya bahan bakar fosil, adopsi mobil listrik secara global mengalami pertumbuhan yang signifikan dari tahun ke tahun. Pemerintah di berbagai negara bahkan sudah mulai mendorong penggunaan EV melalui berbagai kebijakan, insentif, dan pengembangan infrastruktur pendukung seperti stasiun pengisian daya (Dudic, 2024; Feng et al., 2024). Namun, meskipun potensi EV sangat besar, pengembangan ini juga diiringi dengan sejumlah tantangan yang perlu mendapat perhatian.

Salah satu tantangan utama dalam penggunaan mobil listrik adalah fenomena *range anxiety*, yaitu kekhawatiran pengguna bahwa baterai akan habis sebelum sampai tujuan (Chakraborty et al., 2022; Rainieri et al., 2023; Shrestha et al., 2022). Kondisi ini merupakan isu psikologis yang cukup dominan di kalangan pengguna EV, terutama bagi mereka yang sering melakukan perjalanan jauh atau berada di wilayah dengan infrastruktur pengisian daya yang terbatas. Masalah ini tidak hanya memengaruhi kenyamanan dan kepercayaan pengguna terhadap kendaraan listrik, tetapi juga menjadi kendala yang cukup berarti dalam proses adopsi EV secara luas di masyarakat.

Range anxiety menghambat kepercayaan konsumen untuk beralih dari kendaraan berbahan bakar fosil konvensional ke kendaraan listrik, yang dinilai kurang fleksibel dalam hal jangkauan dan ketersediaan energi (Powell & Johnson, 2024). Kekhawatiran ini sering dipicu oleh ketidakakuratan estimasi jarak tempuh yang tersisa pada kendaraan. Estimasi ini biasanya ditampilkan melalui sistem onboard kendaraan, tetapi dalam banyak kasus, informasi yang diberikan tidak cukup akurat atau *real-time*.

Hal ini disebabkan karena sistem estimasi sering kali tidak memperhitungkan secara detail dan dinamis berbagai variabel penting yang memengaruhi konsumsi energi. Faktor-faktor seperti kondisi jalan (menanjak atau menurun), suhu sekitar (terutama suhu ekstrem), gaya dan kebiasaan mengemudi (misalnya akselerasi mendadak atau pengereman intensif), dan pola penggunaan kendaraan sehari-hari semuanya memiliki dampak besar pada masa pakai baterai dan jarak tempuh berkendara sebenarnya (Abuqila et al., 2025; Al-Wreikat et al., 2021; Yang et al., 2023).

Untuk mengatasi permasalahan *range anxiety* yang dipicu oleh estimasi jarak tempuh yang tidak akurat, penggunaan pendekatan *Data Science* dan model algoritma *Machine Learning* (ML) merupakan solusi yang menjanjikan (Mondal et al., 2024; Yaghoubi et al., 2024). Teknologi ini

memungkinkan pengolahan data historis penggunaan kendaraan listrik secara mendalam, sehingga dapat menghasilkan model prediktif yang lebih akurat dan adaptif dalam memperkirakan jarak tempuh kendaraan secara *real-time* (Sang et al., 2024; Zhang et al., 2024a).

Dengan memanfaatkan berbagai fitur seperti riwayat perjalanan, kebiasaan berkendara, kondisi jalan, suhu sekitar, dan pola penggunaan harian, model ML dapat mempelajari pola tersembunyi dalam data yang sulit ditangkap oleh sistem estimasi konvensional. Dengan demikian, estimasi jarak tempuh yang dihasilkan menjadi lebih personal, dinamis, dan sesuai dengan kondisi kendaraan dan lingkungannya yang sebenarnya. Peningkatan akurasi ini diharapkan dapat mengurangi kekhawatiran pengguna dan mendorong rasa percaya diri yang lebih besar dalam penggunaan kendaraan listrik dalam kehidupan sehari-hari.

Berbagai penelitian sebelumnya telah mencoba memprediksi jarak tempuh kendaraan listrik menggunakan algoritma *Machine Learning* (Hasib et al., 2021; Hussain et al., 2025). Meskipun pendekatan ini menunjukkan potensi dalam meningkatkan akurasi estimasi, kebanyakan studi tersebut menerapkan satu algoritma prediksi secara langsung tanpa melalui proses pemilihan fitur yang sistematis. Akibatnya, model yang dihasilkan berisiko mengandung fitur yang kurang relevan atau redundan, yang dapat mengurangi stabilitas model, memperbesar risiko *overfitting*, serta membatasi kemampuan interpretasi terhadap kontribusi masing-masing variabel input. Padahal, identifikasi fitur yang paling berpengaruh sangat penting untuk menghasilkan model prediktif yang bukan hanya akurat tetapi juga dapat diinterpretasikan secara praktis oleh pengguna dan pengembang sistem.

Dengan demikian, terdapat *gap penelitian* yang jelas, yakni kurangnya pendekatan prediktif untuk jarak tempuh EV yang menggabungkan mekanisme pemilihan fitur sistematis dengan model prediksi yang kuat. Kesenjangan ini membuka peluang untuk mengeksplorasi integrasi proses seleksi fitur yang terstruktur dengan teknik prediksi canggih untuk menghasilkan estimasi jarak tempuh yang lebih stabil, akurat, dan adaptif terhadap kondisi penggunaan aktual.

Penelitian ini mengusulkan pendekatan gabungan antara *Random Forest* dan *Xtreme Gradient Boosting (XGBoost)*. Algoritma *Random Forest* akan digunakan untuk memilih fitur-fitur penting dari data, yaitu mengidentifikasi variabel-variabel yang paling mempengaruhi estimasi jarak tempuh. Sementara itu, *XGBoost* digunakan sebagai

model prediksi utama karena kemampuannya untuk membangun pohon keputusan secara berurutan dengan teknik boosting, yang memungkinkan peningkatan akurasi secara iteratif (Fatima et al., 2023). *XGBoost* juga mampu menangani data tabular dengan struktur non-linier secara efisien dan memiliki mekanisme regularisasi bawaan untuk mencegah *overfitting* (Gu et al., 2025; Park & Baek, 2024). Dengan menggabungkan pemilihan fitur dari *Random Forest* dan kekuatan prediksi *XGBoost*, model yang dihasilkan diharapkan mampu memberikan estimasi jarak tempuh mobil listrik yang lebih akurat, stabil, dan interpretable baik secara komputasional maupun praktis (Naseri et al., 2023; Zhang et al., 2024b).

Studi ini bertujuan untuk meningkatkan akurasi prediksi jarak tempuh harian pada mobil listrik menggunakan pendekatan berbasis data dan *Machine Learning*, untuk mengurangi kecemasan pengguna dan meningkatkan kepercayaan terhadap kendaraan listrik.

Secara teoritis, studi ini mengacu pada teori dasar dalam penambangan data, *Machine Learning*, dan studi sebelumnya yang terkait dengan model estimasi energi dan jarak tempuh pada kendaraan listrik. Kontribusi baru dari studi ini terletak pada integrasi proses pemilihan fitur menggunakan *Random Forest* dengan model prediksi berbasis *XGBoost*, pemanfaatan data penggunaan aktual harian yang mempertimbangkan faktor kontekstual pengguna dan lingkungan, dan penerapan pendekatan prediktif yang berorientasi pada solusi kecemasan jarak tempuh dalam kerangka pengembangan sistem manajemen energi berbasis mobilitas cerdas.

Manfaat dan harapan dari penelitian ini adalah terciptanya model prediksi yang mampu memberikan estimasi kebiasaan pengguna yang lebih akurat dan adaptif, yang pada akhirnya dapat mendukung pengembangan sistem manajemen energi kendaraan listrik dan meningkatkan pengalaman pengguna dalam berkendara.

BAHAN DAN METODE

Penelitian ini menggunakan metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang merupakan pendekatan standar dalam proses analisis data dan pengembangan model prediktif berbasis *Machine Learning* (Elkabalawy et al., 2024; Schröer et al., 2021a; Studer et al., 2021). Metodologi ini dipilih karena sangat cocok diaplikasikan pada penelitian berbasis data tabular yang bertujuan untuk menghasilkan prediksi atau klasifikasi berdasarkan fitur teknis (Saltz, 2021).

Business Understanding

Permasalahan utama yang ingin dipecahkan dalam penelitian ini adalah *range anxiety* atau kekhawatiran pengguna kendaraan listrik terhadap ketidakpastian estimasi jarak tempuh. Tujuan dari penelitian ini adalah untuk membangun model prediksi jarak tempuh kendaraan listrik berdasarkan data spesifikasi teknis kendaraan, guna meningkatkan akurasi estimasi dan kepercayaan pengguna terhadap kendaraan listrik.

Data Understanding

Dataset yang digunakan diperoleh dari *open source kaggle* dan terdiri dari total 478 entri dengan 22 atribut teknis, termasuk kapasitas baterai (kWh), efisiensi energi (Wh/km), kecepatan maksimum (km/jam), dimensi kendaraan (mm), akselerasi, dan jenis *drivetrain* (Urvish Ahir, 2025). Pada tahap ini, eksplorasi awal dilakukan untuk memahami distribusi data, jenis variabel, dan deteksi nilai yang hilang.

Tabel 1. Deskripsi Data

Attributes	Description
Brand and Model	Manufacturer and specific nameplate of the EV.
Car Body Type	Classification such as hatchback, SUV, sedan, etc.
Segment	Vehicle segment (e.g., compact, midsize, executive).
Battery Capacity (kWh)	The gross energy capacity of the battery.
Number of Cells and Battery Type	Technical battery information, where available.
Efficiency (Wh/km)	Power consumption rate of the vehicle.
Range (km)	Estimated driving range on a full charge.
Fast Charging Power (kW)	Maximum supported DC fast-charging power.
Fast Charge Port Type	Connector standard (e.g., CCS, CHAdeMO).
Top Speed (km/h)	Maximum speed of the vehicle.
0–100 km/h Acceleration (s)	Time to reach 100 km/h from a standstill.
Torque (Nm)	Maximum torque output, where available.
Towing Capacity (kg)	Ability to tow loads, provided where applicable.
Cargo Volume (L)	Luggage space, sometimes approximate or expressed in alternative units.
Seats	Total seating capacity.
Length, Width, Height (mm)	Physical footprint of the vehicle.
Drivetrain	Powertrain configuration (e.g., AWD, RWD, FWD).
Source URL	Reference link for each car (used in scraping).

Sumber: (Urvish Ahir, 2025)

Data Preparation

Tahap ini melibatkan beberapa proses penting, termasuk:

1. Imputasi nilai yang hilang, menggunakan mean untuk data numerik dan modus untuk data kategorikal.
2. Pengodean fitur kategorikal dilakukan menggunakan *Label Encoding* untuk mengidentifikasi data kategorikal dan numerik.
3. Pemilihan fitur menggunakan *Random Forest Regressor* untuk menentukan fitur yang paling berpengaruh pada target *range_km*.
4. Pembagian data: 80% untuk pelatihan dan 20% untuk pengujian model, hal tersebut merupakan praktik umum dalam machine learning untuk dataset berukuran menengah, sehingga mampu menjaga stabilitas pelatihan dan realibilitas evaluasi model (Elshewey et al., 2025; Hossain, 2025).

Modeling

Model prediktif dibangun menggunakan algoritma *XGBoost Regressor*, algoritma *ensemble* berbasis *boosting* yang dikenal efektif untuk data tabular. Model dilatih menggunakan 10 fitur terbaik yang dipilih dari *Random Forest*, dengan parameter default dan *random_state* untuk replikasi hasil.

Evaluation

Model yang telah dilatih kemudian dievaluasi menggunakan beberapa metrik kuantitatif yaitu *Mean Absolute Error* (MAE) untuk mengukur rata-rata kesalahan absolut, *Root Mean Squared Error* (RMSE) untuk menghitung rata-rata kesalahan kuadrat yang sensitif terhadap *outlier*, dan *R-squared* (R^2) yang merepresentasikan proporsi variasi variabel target yang dapat dijelaskan oleh model.

Selain evaluasi numerik, analisis visual juga dilakukan untuk memperkuat interpretasi hasil model. Visualisasi tersebut meliputi *scatter plot* antara nilai aktual dan prediksi untuk melihat kedekatan hasil prediksi dengan data aktual, histogram distribusi residual untuk memahami sebaran galat prediksi, grafik kepentingan fitur untuk mengidentifikasi kontribusi relatif setiap fitur dalam proses prediksi, dan heatmap korelasi antar fitur numerik yang digunakan untuk mengamati hubungan linier antar variabel numerik dalam dataset. Analisis visual ini berperan penting dalam menilai akurasi dan interpretabilitas model secara keseluruhan.

Metodologi CRISP-DM memastikan bahwa penelitian dilakukan secara sistematis, mulai dari memahami masalah hingga memvalidasi hasil, sehingga mendukung kredibilitas model yang dihasilkan dan memungkinkan replikasi dan pengembangan lebih lanjut (Schröer et al., 2021b).

Deployment

Model *XGBoost* yang dihasilkan dalam studi ini dapat dianggap sebagai bukti konsep untuk sistem prediksi jarak tempuh mobil listrik berdasarkan data spesifikasi teknis. Dengan akurasi tinggi dan kemampuan generalisasi yang baik, model ini berpotensi untuk diintegrasikan ke dalam sistem informasi kendaraan, aplikasi konfigurator mobil listrik, atau dasbor rekomendasi pembelian kendaraan berdasarkan kebutuhan harian pengguna.

Selain itu, model ini juga dapat menjadi dasar pengembangan sistem pendukung keputusan bagi produsen kendaraan dalam merancang produk yang sesuai dengan kebutuhan pasar. Tahapan *deployment* dalam penelitian ini berfokus pada dokumentasi model, visualisasi hasil, dan pemetaan arah implementasi model masa mendatang dalam skenario dunia nyata, baik yang terintegrasi dalam sistem berbasis web maupun sebagai modul dalam perangkat lunak pengujian kendaraan.

Dengan demikian, meskipun penelitian ini tidak melibatkan pengembangan perangkat lunak atau platform aplikasi secara langsung, pendekatan yang diambil telah mempertimbangkan relevansi praktis dari model yang dibangun dan membuka ruang untuk adopsi lebih lanjut dan pengembangan sistem dalam fase pasca penelitian.

Encoding Data Kategorikal

Dataset kendaraan listrik pada penelitian ini memiliki beberapa atribut kategorikal, seperti model, brand yang tidak dapat diproses secara langsung oleh algoritma machine learning berbasis pohon keputusan. Oleh karena itu, penelitian ini menerapkan *Label Encoding* untuk mentransformasikan fitur kategorikal menjadi representasi numerik berurutan (Sree et al., 2021).

Metode tersebut dipilih karena model yang digunakan dalam penelitian ini yaitu *XGBoost* dan *Random Forest*, merupakan algoritma berbasis pohon keputusan yang tidak mengasumsikan hubungan ordinal atau linear antar nilai numerik input (Elshewey et al., 2025). Dengan demikian transformasi numerik melalui label encoding tidak menyebabkan bias struktural terhadap proses pembentukan pohon keputusan.

Untuk mengendalikan potensi redundansi fitur hasil encoding, penelitian ini mengkombinasikan *Label Encoding* dengan pemilihan fitur berbasis *Random Forest*, sehingga hanya fitur yang memiliki kontribusi signifikan terhadap prediksi jarak tempuh (*range_km*) yang digunakan pada pemodelan lanjutan.

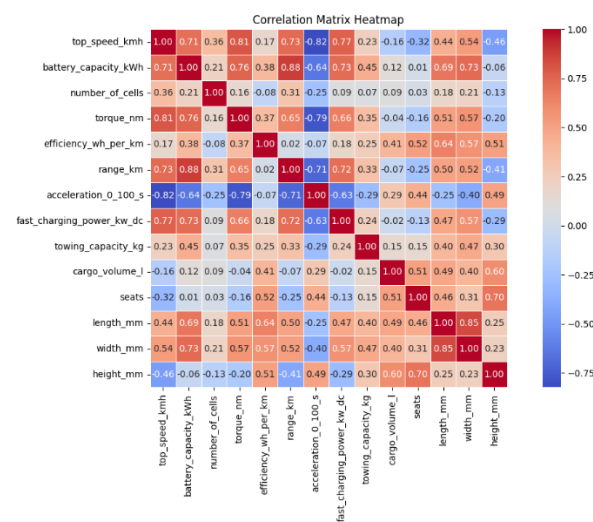
HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini terdiri dari 478 entri yang mewakili spesifikasi teknis berbagai model mobil listrik dari berbagai merek. Setiap entri memiliki 22 atribut, termasuk informasi seperti kapasitas baterai (kWh), kecepatan maksimum (km/jam), efisiensi energi (Wh/km), dimensi kendaraan (mm), dan parameter teknis lainnya.

Setelah pra-pemrosesan dan imputasi nilai yang hilang, proses mengubah data kategorikal menjadi numerik dilakukan dengan menerapkan metode *Label Encoding*, hal ini dikarenakan dataset memiliki atribut kategorikal sehingga perlu disesuaikan menjadi numerik agar dapat diproses. Penerapan *Label Encoding* memungkinkan integrasi fitur kategorikal ke dalam model prediktif tanpa meningkatkan dimensi fitur secara signifikan.

Hasil evaluasi menunjukkan bahwa penggunaan *Label Encoding* tidak menurunkan kinerja model *XGBoost*, yang tercermin dari nilai nilai R^2 yang tinggi. Hal ini mengindikasikan bahwa model mampu mempelajari pola relasi antar kategori dan variabel target tanpa terpengaruh oleh representasi numerik kategori, dengan dikombinasikan bersama proses pemilihan fitur *Random Forest*, *Label Encoding* berkontribusi pada stabilitas model dan menjaga kemampuan generalisasi, sekaligus mempertahankan interpretabilitas hasil prediksi.

proses pemilihan fitur dilakukan menggunakan algoritma *Random Forest*. Fitur dengan tingkat kepentingan tertinggi dipilih untuk melatih model *XGBoost*. Hal ini bertujuan untuk mengurangi kompleksitas model dan fokus pada variabel yang paling memengaruhi jarak tempuh.



Sumber: (Hasil Penelitian, 2025)
Gambar 1. *Heatmap* Matriks Korelasi

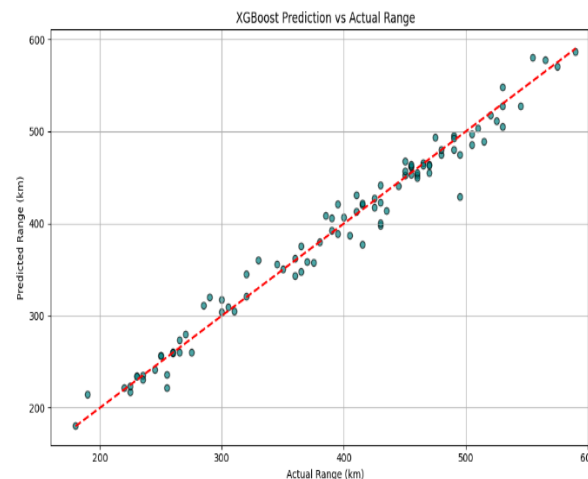
Heatmap pada gambar 1 menunjukkan nilai korelasi *Pearson* antara fitur numerik. Dapat dilihat bahwa *battery_capacity_kWh* berkorelasi positif dengan *range_km*, sedangkan *efficiency_wh_per_km* berkorelasi negatif, yang sejalan dengan ekspektasi teoritis.

Tabel 2. Top 10 Korelasi Variabel

No	Fitur	Korelativitas
1	battery_capacity_kWh	0.799
2	height_mm	0.061
3	efficiency_wh_per_km	0.023
4	Segment	0.02
5	fast_charging_power_kw_dc	0.02
6	top_speed_kmh	0.019
7	acceleration_0_100_s	0.017
8	cargo_volume_l	0.008
9	width_mm	0.006
10	number_of_cells	0.005

Sumber: (Hasil Penelitian, 2025)

Random Forest menunjukkan bahwa *battery_capacity_kWh*, *height_mm*, dan *efficiency_wh_per_km* memiliki efek signifikan pada jarak tempuh yang diprediksi. Temuan ini relevan untuk sistem manajemen baterai kendaraan listrik. Visualisasi perbandingan antara nilai aktual dan prediksi dari model *XGBoost* ditunjukkan pada gambar 2.



Sumber: (Hasil Penelitian, 2025)
Gambar 2. Visualisasi Prediksi vs Data Aktual

Berdasarkan gambar 2. Visualiasi Prediksi vs Data Aktual, menunjukkan distribusi titik prediksi terhadap nilai aktual dengan garis diagonal sebagai representasi prediksi yang sempurna. Model *XGBoost* yang dilatih menggunakan 10 fitur terbaik menunjukkan kinerja prediksi yang sangat baik.

Model *XGBoost* pada tahap awal di latih menggunakan parameter default sebagai baseline evaluasi. Selanjutnya, dilakukan *Hyperparameter tuning* menggunakan pendekatan *Grid Search*

dengan *k-fold cross-validation* untuk memperoleh konfigurasi parameter optimal. Parameter yang dioptimalkan meliputi *max_depth*, *learning_rate*, *n_estimator*, dan *subsample*. Proses tuning dilakukan setelah tahap *encoding* dan pemilihan fitur, sehingga model yang dihasilkan mencerminkan pengaruh kombinasi *Label Encoding*, *Featur Selection*, dan konfigurasi parameter yang optimal.

Optimalisasi model *XGBoost* dalam memprediksi jangkauan kendaraan listrik (*range_km*) digunakan pendekatan *Grid Search*, parameter didefinisikan dengan rentang nilai *max_depth* [3, 5, 7], *learning_rate* [0.01, 0.1, 0.2], *n_estimator* [100, 200, 300], *subsample* [0.7, 0.8, 0.9].

Grid Search diinisialisasi dengan *XGBoost* dan dikonfigurasi untuk menggunakan *neg_mean_absolute_error* sebagai metrik penilaian dengan *cross-validation 3-fold*. Pencarian dilakukan pada data pelatihan (*X-train*, *y-train*) untuk secara sistematis mengeksplorasi semua kombinasi *Hyperparameter*.

Setelah dilakukan *fitting Grid Search*, *hyperparameter* optimal diidentifikasi {'*learning_rate*': 0.2, '*max_depth*': 3, '*n_estimators*': 300, '*subsample*': 0.8}. Model baru kemudian dilatih menggunakan parameter terbaik yang telah didapatkan. Performa model dievaluasi dan dibandingkan dengan model awal.

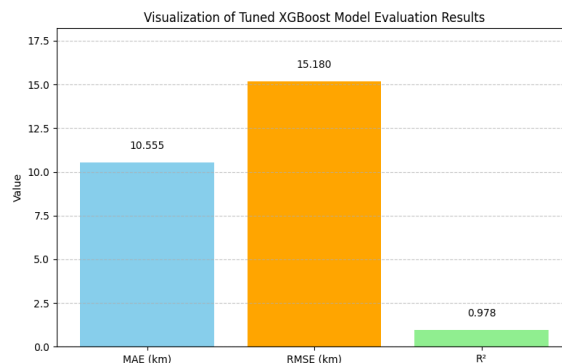
Tabel 3. Perbandingan Evaluasi Model

Matrix	Model awal	Model yang dioptimalkan
MAE	11.382	10.555
RMSE	15.685	15.180
R-Squared (R^2)	0.977	0.978

Sumber: (Hasil Penelitian, 2025)

Proses pengoptimalan *Hyperparameter* menghasilkan sedikit peningkatan performa model, dengan penurunan *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE), serta nilai *R-squared* yang sedikit lebih tinggi, menunjukkan kesesuaian yang lebih baik dengan data. Hal ini menunjukkan bahwa *Hyperparameter* yang dioptimalkan berkontribusi pada prediksi jangkauan kendaraan listrik (*range_km*) yang lebih akurat dan kuat. Matriks evaluasi model pada data uji menunjukkan:

- *Mean Absolute Error* (MAE): 10.555
- *Root Mean Squared Error* (RMSE): 15.180
- *R-squared* (R^2): 0.978



Sumber: (Hasil Penelitian, 2025)

Gambar 3. Visualisasi Hasil Evaluasi Model *XGBoost*

Visual yang terdapat pada Gambar 3 menunjukkan hasil evaluasi model prediktif *XGBoost* dengan menggunakan tiga metrik utama, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-Squared* (R^2). Nilai MAE sebesar 10.555 menunjukkan bahwa rata-rata kesalahan prediksi model terhadap jarak tempuh aktual mobil listrik relatif kecil. Sementara itu, nilai RMSE sebesar 15.180 menunjukkan deviasi prediksi yang masih dalam batas wajar, meskipun RMSE secara alami lebih sensitif terhadap *outlier* daripada MAE. Nilai R^2 yang sangat tinggi yaitu 0.978 menegaskan bahwa model mampu menjelaskan 97,8% variasi pada data target *range_km*.

Hal ini menunjukkan bahwa fitur teknis kendaraan listrik memiliki hubungan yang kuat dengan jarak tempuh kendaraan yang sebenarnya. Visualisasi ini juga memperjelas bahwa model yang dibangun memiliki kinerja yang kuat dan stabil, serta mampu memberikan estimasi jarak tempuh dengan tingkat akurasi yang tinggi. Hal ini memperkuat kesimpulan bahwa kombinasi metode pemilihan fitur *Random Forest* dan pemodelan dengan *XGBoost* efektif dalam konteks data spesifikasi teknis kendaraan listrik.

Secara khusus, fitur *battery_capacity_kWh* terbukti menjadi penentu utama, sejalan dengan ekspektasi logis bahwa kapasitas baterai berbanding lurus dengan energi yang tersedia untuk jarak tempuh. Selain itu, dimensi kendaraan seperti *height_mm* dan efisiensi konsumsi energi (*efficiency_wh_per_km*) juga memberikan kontribusi yang signifikan.

Fitur-fitur ini memberikan wawasan penting bahwa karakteristik fisik dan efisiensi aerodinamis berperan dalam kinerja kendaraan. Hasil ini menunjukkan bahwa kombinasi *Random Forest* untuk pemilihan fitur dan *XGBoost* untuk regresi sangat efektif pada data tabular EV. Fitur teknis kendaraan mengandung cukup informasi untuk

membuat prediksi tentang kinerja fungsional seperti jarak tempuh.

Dari sisi performa model terbaik, *XGBoost* pada penelitian ini memiliki nilai R^2 0.978, lebih tinggi dibandingkan penelitian yang dilakukan Hussain, Izhar dkk (Hussain et al., 2025) performa model terbaik yang dimilikinya yaitu 0.9592 dan penelitian Hasib dkk (Hasib et al., 2021) yang menunjukkan performa melalui R^2 sebesar 0.973 menggunakan Multiple Linear Regression (MLR). Hal ini menunjukkan bahwa kombinasi seleksi fitur dan pemodelan berbasis boosting efektif untuk data spesifikasi teknis EV.

Penelitian ini membuka jalan bagi integrasi data teknis dengan data pola berkendara di dunia nyata, seperti GPS, suhu, dan kebiasaan pengisian daya untuk membangun model hibrida prediktif. Pengembangan sistem rekomendasi berbasis spesifikasi teknis juga merupakan arah yang menarik untuk penelitian lebih lanjut di bidang kendaraan listrik dan kecerdasan buatan.

KESIMPULAN

Studi ini berhasil membangun model prediktif untuk estimasi jarak tempuh (*range_km*) mobil listrik (*Electric Vehicle/EV*) berdasarkan spesifikasi teknis kendaraan menggunakan kombinasi algoritma *Random Forest* dan *XGBoost*. Hasil evaluasi menunjukkan bahwa model *XGBoost* yang dilatih pada 10 fitur terpenting menggunakan pendekatan *Label Encoding* dan *Hyperparameter Tuning* model menghasilkan kinerja prediksi yang sangat baik, dengan nilai R^2 sebesar 0,978, MAE sebesar 10,555, dan RMSE sebesar 15,180. Fitur *battery_capacity_kWh*, *efficiency_wh_per_km*, dan *height_mm* dicatat sebagai faktor dominan yang memengaruhi estimasi jarak tempuh.

Kontribusi utama dari studi ini terletak pada pendekatan metodologis yang menggabungkan pemilihan fitur berbasis *Random Forest* dan pemodelan prediktif dengan *XGBoost*, yang telah terbukti efektif dalam memodelkan data spesifikasi teknis non-sekuensial. Pendekatan ini juga memberikan kontribusi praktis dalam membantu produsen kendaraan listrik memberikan estimasi jarak tempuh yang lebih akurat kepada konsumen, serta menjadi solusi potensial untuk mengurangi kecemasan jarak tempuh.

Implikasi dari penelitian ini adalah data spesifikasi teknis kendaraan dapat dimanfaatkan secara optimal untuk membangun sistem prediksi jarak tempuh yang akurat dan efisien. Hasil ini dapat mendukung pengembangan sistem informasi kendaraan listrik dan membuka peluang integrasi dalam aplikasi berbasis rekomendasi kendaraan. Akan tetapi, penelitian ini memiliki beberapa

keterbatasan, antara lain tidak adanya data penggunaan di dunia nyata seperti pola berkendara, kondisi cuaca, dan medan jalan, yang juga memengaruhi konsumsi energi kendaraan. Selain itu, data yang digunakan merupakan data statis tanpa dimensi temporal, sehingga model tidak dapat menangkap dinamika penggunaan kendaraan yang sebenarnya.

REFERENSI

- Abuqila, M., Nassar, Y., & Nyasapoh, M. (2025). Estimation of the Storage Capacity of Electric Vehicle Batteries under Real Weather and Drive-mode Conditions: A Case Study. *Wadi Alshatti University Journal of Pure and Applied Sciences*, 58–71.
- Al-Ghaili, A. M., Kasim, H., Aris, H., & Al-Hada, N. M. (2022). Can electric vehicles be an alternative for traditional fossil-fuel cars with the help of renewable energy sources towards energy sustainability achievement? *Energy Informatics*, 5(Suppl 4), 60.
- Al-Wreikat, Y., Serrano, C., & Sodr , J. R. (2021). Driving behaviour and trip condition effects on the energy consumption of an electric vehicle under real-world driving. *Applied Energy*, 297, 117096.
- Chakraborty, P., Parker, R., Hoque, T., Cruz, J., Du, L., Wang, S., & Bhunia, S. (2022). Addressing the range anxiety of battery electric vehicles with charging en route. *Scientific Reports*, 12(1), 5588.
- Dudic, B. (2024). Global Electric Car Market. *International Conference "New Technologies, Development and Applications,"* 158–164.
- Elkabalawy, M., Al-Sakkaf, A., Mohammed Abdelkader, E., & Alfalah, G. (2024). CRISP-DM-Based Data-Driven Approach for Building Energy Prediction Utilizing Indoor and Environmental Factors. *Sustainability*, 16(17), 7249.
- Elshewey, A. M., Osman, A. M., Youssef, R. Y., El-Bakry, H. M., & Hassan, S. A. Z. (2025). Enhancing forest fires classification using a hybrid convolutional and BiLSTM deep learning model. *Modeling Earth Systems and Environment*, 11(5), 379.
- Fatima, S., Hussain, A., Amir, S. Bin, Ahmed, S. H., & Aslam, S. M. H. (2023). XGBoost and random forest algorithms: an in depth analysis. *Pakistan Journal of Scientific Research*, 3(1), 26–31.
- Feng, J., Yao, Y., Liu, Zhenfeng, & Liu, Zhenling. (2024). Electric vehicle charging stations' installing strategies: Considering government subsidies. *Applied Energy*, 370, 123552.

- Gu, M., Kang, S., Xu, Z., Lin, L., & Zhang, Z. (2025). AE-XGBoost: A Novel Approach for Machine Tool Machining Size Prediction Combining XGBoost, AE and SHAP. *Mathematics*, 13(5), 835.
- Hasib, S. A., Saha, D. K., Islam, S., Tanvir, M., & Alam, M. S. (2021). Driving range prediction of electric vehicles: A machine learning approach. *2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 1–6.
- Hossain, M. R. (2025). Predicting customer churn in telecommunications with machine learning models. *Asian Journal of Research in Computer Science*, 18(1), 53–66.
- Hussain, I., Ching, K. B., Uttraphan, C., Tay, K. G., & Noor, A. (2025). Evaluating machine learning algorithms for energy consumption prediction in electric vehicles: A comparative study. *Scientific Reports*, 15(1), 16124.
- Mondal, S., Ramesh, H. N., Ma, X., Cheng, S., Angeles, T., & Sun, S. (2024). *Reducing electric vehicle range anxiety with machine learning models incorporating human behavior*.
- Naseri, H., Waygood, E. O. D., Wang, B., & Patterson, Z. (2023). Interpretable machine learning approach to predicting electric vehicle buying decisions. *Transportation Research Record*, 2677(12), 704–717.
- Park, J.-H., & Baek, S. H. (2024). Two-Step Methodology for Statistical Anomaly Detection and Prediction Using XGBoost Regression in Blower Motor Vibration Time Series Data. *International Journal of Industrial Engineering: Theory, Applications and Practice*, 31(4).
- Powell, B., & Johnson, C. (2024). *Impact of Electric Vehicle Charging Station Reliability, Resilience, and Location on Electric Vehicle Adoption*. National Renewable Energy Laboratory (NREL), Golden, CO (United States).
- Rainieri, G., Buizza, C., & Ghilardi, A. (2023). The psychological, human factors and socio-technical contribution: A systematic review towards range anxiety of battery electric vehicles' drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 99, 52–70.
- Saltz, J. S. (2021). CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps. *2021 IEEE International Conference on Big Data (Big Data)*, 2337–2344. <https://doi.org/10.1109/BigData52589.2021.9671634>
- Sang, V. T. D., Duong, Q. H., Zhou, L., & Arranz, C. F. A. (2024). Electric vehicle battery technologies and capacity prediction: a comprehensive literature review of trends and influencing factors. *Batteries*, 10(12), 451.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021a). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/https://doi.org/10.1016/j.procs.2021.01.199>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021b). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Shrestha, S., Baral, B., Shah, M., Chitrakar, S., & Shrestha, B. P. (2022). Measures to resolve range anxiety in electric vehicle users. *International Journal of Low-Carbon Technologies*, 17, 1186–1206.
- Sinha, P., Paul, K., Deb, S., & Sachan, S. (2023). Comprehensive review based on the impact of integrating electric vehicle and renewable energy sources to the grid. *Energies*, 16(6), 2924.
- Sree, K. S., Karthik, J., Niharika, C., Srinivas, P., Ravinder, N., & Prasad, C. (2021). Optimized conversion of categorical and numerical features in machine learning models. *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 294–299.
- Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K.-R. (2021). Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology. *Machine Learning and Knowledge Extraction*, 3(2), 392–413.
- Sudjoko, C., Sasongko, N. A., Utami, I., & Maghfuri, A. (2021). Utilization of electric vehicles as an energy alternative to reduce carbon emissions. *IOP Conference Series: Earth and Environmental Science*, 926(1), 012094.
- Urvish Ahir. (2025). *Electric Vehicle Specs Dataset (2025)*. Kaggle. <https://www.kaggle.com/datasets/urvishahir/electric-vehicle-specifications-dataset-2025>
- Yaghoubi, Elaheh, Yaghoubi, Elnaz, Khamees, A., Razmi, D., & Lu, T. (2024). A systematic review and meta-analysis of machine learning, deep learning, and ensemble learning approaches in predicting EV charging behavior. *Engineering Applications of Artificial Intelligence*, 135, 108789.
- Yang, D., Liu, H., Li, M., & Xu, H. (2023). Data-driven analysis of battery electric vehicle energy

consumption under real-world temperature conditions. *Journal of Energy Storage*, 72, 108590.

Zhang, W., Fang, X., & Sun, C. (2023). The alternative path for fossil oil: electric vehicles or hydrogen fuel cell vehicles? *Journal of Environmental Management*, 341, 118019.

Zhang, X., Zhang, Z., Liu, Y., Xu, Z., & Qu, X. (2024a). A review of machine learning approaches for electric vehicle energy consumption

modelling in urban transportation. *Renewable Energy*, 234, 121243.

Zhang, X., Zhang, Z., Liu, Y., Xu, Z., & Qu, X. (2024b). A review of machine learning approaches for electric vehicle energy consumption modelling in urban transportation. *Renewable Energy*, 234, 121243.