

p-ISSN: 2774-4574 ; E-ISSN: 363-4582
TRILOGI, 5(3), Juli-Sep 2024 (456-467)
©2020 Lembaga Penerbitan, Penelitian,
dan Pengabdian kepada Masyarakat (LP3M)
Universitas Nurul Jadid Paiton Probolinggo
DOI: [10.33650/trilogi.v5i3.9000](https://doi.org/10.33650/trilogi.v5i3.9000)

JURNAL TRILOGI
Ilmu Teknologi, Kesehatan, dan Humaniora

Improve Metode Lightgbm untuk Prediksi Harga Mobil Bekas Menggunakan Hyper-Parameter Tuning

Moch. Aqil Aulady

Universitas Nurul Jadid, Indonesia
mochaqilaulady13@gmail.com

Ahmad Hudawi AS.

Universitas Nurul Jadid, Indonesia
ahmad.hudawi@unuja.ac.id

Zainal Arifin

Universitas Nurul Jadid, Indonesia
zainal@unuja.ac.id

Abstract

This study aims to predict used car prices using the LightGBM method and hyperparameter tuning techniques in the context of data science. The analysis process includes collecting historical data on used cars, preprocessing the data to clean and encode variables, and splitting the data into training and testing sets. The LightGBM model was trained and optimized through hyperparameter tuning using GridSearchCV to improve model performance. The model was evaluated using metrics such as Mean Squared Error (MSE) and R-squared. The results indicate that the well-optimized LightGBM model can accurately predict used car prices with high accuracy. The low MSE value (35207938112.028404) and high R-squared value (0.9462871489515565) demonstrate the model's excellent predictive quality. This research provides deeper insights into the factors influencing used car prices and contributes to the development of effective and reliable predictive models.

Keywords: LightGBM; hyperparameter tuning; price prediction; used cars; GridSearchCV.

Abstrak

Penelitian ini bertujuan untuk memprediksi harga mobil bekas menggunakan metode LightGBM dan teknik tuning hyperparameter dalam konteks ilmu data. Proses analisis meliputi pengumpulan data historis mobil bekas, prapemrosesan data untuk membersihkan dan mengonversi variabel, serta membagi data menjadi set pelatihan dan pengujian. Model LightGBM dilatih dan dioptimalkan melalui tuning hyperparameter menggunakan GridSearchCV untuk meningkatkan kinerja model. Evaluasi model dilakukan dengan menggunakan metrik seperti Mean Squared Error (MSE) dan R-squared. Hasil penelitian menunjukkan bahwa model LightGBM yang dioptimalkan dapat memprediksi harga mobil

bekas dengan tingkat akurasi yang tinggi. Nilai MSE yang rendah (35207938112.028404) dan nilai R-squared yang tinggi (0.9462871489515565) menunjukkan bahwa model ini memiliki kualitas prediksi yang sangat baik. Penelitian ini memberikan wawasan mendalam tentang faktor-faktor yang memengaruhi harga mobil bekas serta berperan dalam pengembangan model prediksi yang efektif dan andal.

Katakunci: LightGBM; Hyperparameter Tuning; Prediksi Harga; Mobil Bekas; GridSearchCV.

1 Pendahuluan

Mobil bekas merupakan salah satu aset bergerak yang memiliki nilai signifikan dalam pasar otomotif global. Keberadaan mobil bekas menawarkan berbagai keuntungan, terutama dari segi biaya, yang membuatnya menjadi pilihan utama bagi banyak konsumen. Mobil bekas memiliki karakteristik yang dinamis dan kompleks, yang dapat berdampak pada kenyamanan penggunaannya (Butsianto & Mayangwulan, 2020). Variasi dalam usia dan spesifikasi teknis mobil bekas dapat berdampak pada pengalaman berkendara serta kenyamanan pengguna. Oleh karena itu, memahami faktor-faktor yang mempengaruhi harga mobil bekas sangat penting untuk menilai kendaraan dan membuat keputusan pembelian yang lebih terinformasi.

Harga mobil bekas dipengaruhi oleh berbagai faktor yang saling terkait dan bersifat multidimensi. Faktor-faktor ini tidak hanya menentukan nilai pasar kendaraan tetapi juga berperan dalam daya tarik mobil bekas di pasar. Beberapa faktor utama yang berkontribusi pada penetapan harga mobil bekas meliputi, tahun produksi kendaraan (Kriswantara et al., 2021) berhubungan erat dengan depresiasi kendaraan mobil yang lebih tua biasanya mengalami penurunan nilai yang lebih besar dibandingkan dengan model yang lebih baru. Inovasi teknologi dan perbaikan dalam model terbaru juga mempengaruhi persepsi nilai mobil bekas di pasar, merek kendaraan (Sofyan & Nilmada, 2023) adalah faktor penting lainnya dalam menentukan harga mobil bekas. Merek tertentu sering dikenal dengan kualitas dan daya tahan yang lebih tinggi, yang dapat meningkatkan nilai jual kembali kendaraan tersebut, Tipe kendaraan (Pamungkas et al., 2023) seperti sedan, SUV, *hatchback*, atau tipe lainnya, juga memainkan peran krusial dalam penetapan harga. Setiap tipe kendaraan memiliki pasar dan permintaan yang berbeda, yang mempengaruhi nilai jual kembali, model kendaraan (Nasyuli et al., 2023) yang spesifik sering kali mencakup fitur-fitur unik atau varian

khusus yang dapat mempengaruhi harga. Model dengan fitur tambahan atau varian edisi khusus biasanya memiliki harga yang lebih tinggi dibandingkan dengan model standar, usia kendaraan (Kusuma & Hidayat, 2024) juga merupakan faktor kunci dalam menentukan harga. Kendaraan yang lebih tua biasanya mengalami penurunan nilai yang lebih besar dibandingkan dengan kendaraan yang lebih baru dan jarak tempuh (Surya Negara et al., 2023) kendaraan menjadi indikator penting dalam menentukan harga mobil bekas. Jarak tempuh yang lebih tinggi sering kali menunjukkan bahwa kendaraan telah digunakan secara intensif, yang dapat mempengaruhi kondisi mekanis dan, pada akhirnya, nilai jual kendaraan.

Dalam bidang data *science*, ada berbagai metode yang dapat dimanfaatkan untuk melakukan prediksi (Nugraha, 2021) harga mobil bekas. Metode ini bertujuan untuk memahami dan mengukur hubungan antara berbagai faktor yang mempengaruhi harga, serta mengembangkan model yang dapat memberikan prediksi yang akurat. Beberapa metode yang sering diterapkan dalam analisis prediktif seperti *Regresi Linier* dan *Random Forest* (Radhi et al., 2021). *Regresi Linier* adalah salah satu metode yang paling dasar dan sederhana. Metode ini berfungsi dengan mengasumsikan adanya hubungan linear antara variabel independen dan variabel dependen. Meskipun *Regresi Linier* mudah dipahami dan diterapkan, metode ini memiliki keterbatasan signifikan ketika dihadapkan dengan fitur-fitur *non-linier* dan interaksi kompleks antar fitur. *Regresi Linier* sering kali tidak mampu menangani hubungan yang kompleks antara variabel, sehingga model yang dihasilkan mungkin kurang akurat dalam konteks data yang sangat variatif dan *non-linier* (Hasibuan & Karim, 2022). Sementara itu, metode *Random Forest* merupakan teknik ensemble learning yang dibangun dengan cara membentuk beberapa pohon keputusan dan menggabungkan hasil prediksi dari masing-masing pohon untuk meningkatkan akurasi dan mengurangi risiko overfitting (Pandey et al., 2020).

Karena itu, dalam penelitian ini, dipilih metode *LightGBM* (Febriantoro et al., 2023), (Diantika, 2023), (Sari et al., 2023), (M John et al., 2022), (Tang et al., 2020), (Cao et al., 2023) sebagai pendekatan utama dalam model prediksi harga mobil bekas. *LightGBM* adalah teknik *gradient boosting* yang dirancang untuk menangani dataset besar dengan efisiensi tinggi dan kecepatan pelatihan model (X. Yang et al., 2024), (Zhang et al., 2020), yang sangat cepat. Metode ini terkenal dengan kemampuannya dalam menangani fitur-fitur kompleks dan data yang besar secara efektif, serta dapat menangani data yang tidak seimbang dengan baik. *LightGBM* juga memiliki kemampuan untuk mengoptimalkan berbagai parameter dalam model, yang memungkinkan penyesuaian yang lebih baik terhadap data yang digunakan.

Meskipun *LightGBM* menawarkan banyak keunggulan, efektivitas metode ini sangat bergantung pada pengaturan *hyperparameter* (Alibrahim & Ludwig, 2021) yang tepat. Pengaturan *hyperparameter* yang tidak optimal dapat menyebabkan kinerja model yang kurang baik dan dapat mengakibatkan masalah *overfitting*, di mana model menjadi terlalu kompleks dan tidak mampu generalisasi dengan baik pada data baru. Oleh karena itu, penting untuk melakukan *hyperparameter tuning* secara cermat untuk memastikan bahwa model *LightGBM* dapat meningkatkan kinerja model *LightGBM* dalam memprediksi (X. Yang et al., 2024) harga mobil bekas dengan lebih akurat dan mengurangi risiko *overfitting* pada dataset yang digunakan untuk mencapai performa terbaiknya.

Sebagai solusi, penelitian ini akan fokus pada proses *hyperparameter tuning* yang mendalam pada *LightGBM*. *Hyperparameter tuning* adalah proses yang melibatkan penyesuaian parameter-parameter dalam model untuk meningkatkan akurasi dan kinerja model secara keseluruhan. Dengan melakukan *tuning hyperparameter* secara sistematis, diharapkan model *LightGBM* dapat mencapai akurasi yang lebih tinggi dalam memprediksi harga mobil bekas dan mengurangi risiko *overfitting* pada dataset yang digunakan. Penelitian ini bertujuan untuk mengembangkan model prediksi harga mobil bekas yang tidak hanya akurat tetapi juga optimal dengan memanfaatkan teknik *tuning hyperparameter* pada *LightGBM* (H. Yang et al., 2023), (Zhu et al., 2023). Penelitian ini bertujuan untuk memberikan pemahaman yang lebih mendalam tentang faktor-faktor yang mempengaruhi harga mobil bekas serta untuk mengembangkan metodologi yang

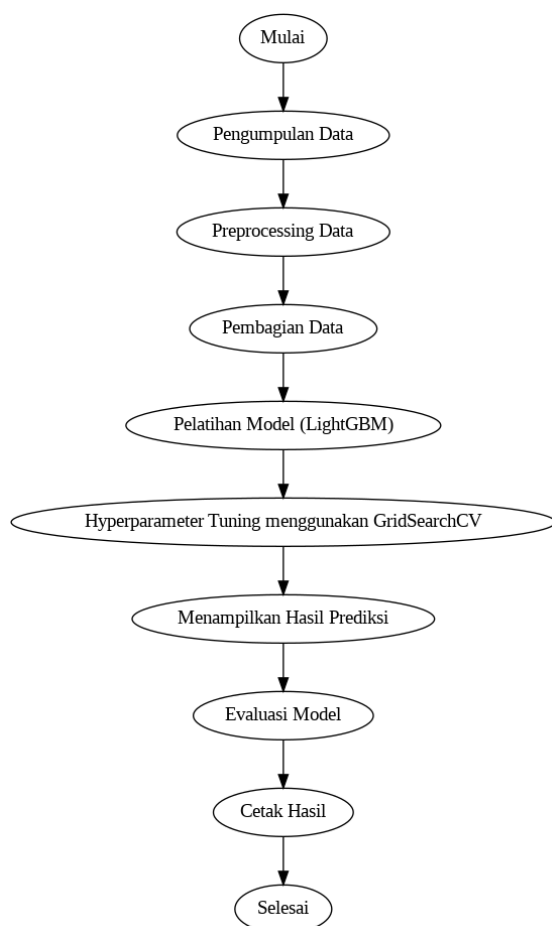
efektif dalam memprediksi harga tersebut. Dengan memanfaatkan metode *LightGBM* dan teknik *hyperparameter tuning*, diharapkan penelitian ini dapat mengatasi keterbatasan yang ada dalam model-model prediksi sebelumnya

2 Metode

Penelitian ini menerapkan metode kuantitatif untuk menguji prediksi harga mobil bekas, dengan menggunakan algoritma *LightGBM*. Metode kuantitatif memungkinkan pengumpulan data secara sistematis dan analisis statistik yang terukur, memastikan pendekatan yang terstruktur dan valid dalam setiap langkah penelitian. Pertama, adalah mengumpulkan data historis harga mobil bekas yang mencakup variabel numerik seperti tahun produksi, jarak tempuh, dan harga jual sebelumnya. Kemudian, dilakukan preprocessing data untuk membersihkan, mengisi nilai yang hilang, normalisasi data, dan mengubah variabel kategorikal menjadi bentuk numerik yang sesuai. Setelah itu, membagi data menjadi set pelatihan dan set pengujian untuk evaluasi model.

Langkah berikutnya adalah melakukan *tuning hyperparameter* dengan teknik *GridSearchCV* untuk mencari kombinasi *hyperparameter* yang dapat meningkatkan kinerja model *LightGBM*. Proses ini sangat penting karena *hyperparameter* yang optimal dapat menghasilkan model yang lebih akurat dan mengurangi risiko *overfitting* pada data pelatihan. Setelah mendapatkan kombinasi *hyperparameter* yang optimal, model *LightGBM* dilatih menggunakan set pelatihan dan kinerja model diukur dengan metrik evaluasi kuantitatif seperti *Mean Squared Error (MSE)* atau *R-squared*. Selanjutnya, validasi model dilakukan pada set pengujian untuk memastikan kinerja yang baik dan mencegah *overfitting*. Validasi dilakukan dengan membandingkan prediksi harga mobil bekas dari model dengan harga sebenarnya pada data pengujian. Dengan mengikuti langkah-langkah ini, diharapkan dapat menghasilkan model prediksi harga mobil bekas yang akurat dan dapat diandalkan, serta mengidentifikasi faktor-faktor yang secara signifikan memengaruhi harga mobil bekas berdasarkan analisis data numerik dan statistik.

Diagram alur berikut ini menggambarkan proses pengumpulan dan pengolahan data secara lebih detail.



Gambar. 1. Skema Analisis Penelitian.

Dalam pengembangan model Machine Learning untuk memprediksi harga mobil bekas menggunakan *LightGBM* dapat dijelaskan sebagai berikut;

1. Mulai:

Titik awal dari proses analisis penelitian.

2. Pengumpulan Data:

Tahap ini mengumpulkan data dari sumber data yang relevan (Nugraha, 2021), (Shudiq et al., 2020) seperti Kaggle.com. Data yang dikumpulkan mencakup informasi tentang mobil bekas seperti model mobil, tahun pembuatan, kondisi mesin, dan harga jual.

3. Pemrosesan Data:

Setelah data terkumpul, langkah selanjutnya adalah melakukan pemrosesan data (Jesika et al., 2023). Proses ini mencakup pembersihan data dari nilai yang hilang atau tidak valid, pengkodean variabel kategorikal

(jika ada), dan normalisasi data numerik agar data siap untuk digunakan dalam pembangunan model.

4. Pemisahan Data:

Setelah data diproses, dilakukan pemisahan data menjadi dua bagian utama: data latih (train data) dan data uji (test data) (Kurniawan et al., 2023). Data latih digunakan untuk melatih model, sementara data uji digunakan untuk menguji kinerja model yang telah dibangun.

5. Pelatihan Model (*LightGBM*):

Model Machine Learning yang digunakan dalam penelitian ini adalah *LightGBM*, sebuah algoritma boosting tree yang efisien dalam menangani dataset besar. Model ini dilatih menggunakan data latih (Duran, 2023) untuk mempelajari pola-pola dan hubungan antar variabel.

6. *Hyperparameter Tuning* menggunakan *GridSearchCV*:

Setelah model dibangun, dilakukan proses tuning pada hyperparameter menggunakan *GridSearchCV*. *GridSearchCV* (Darmawan & Dianta, 2023) adalah metode pencarian sistematis untuk menemukan kombinasi hyperparameter terbaik yang mengoptimalkan kinerja model.

7. Menampilkan Hasil Prediksi:

Setelah model di-tune, dilakukan prediksi harga mobil bekas menggunakan data uji. Hasil prediksi ini kemudian ditampilkan untuk mengevaluasi seberapa baik model dalam memprediksi harga mobil yang sebenarnya.

8. Evaluasi Model:

Performa model dievaluasi menggunakan metrik (Azriel & Darmawan, 2024) evaluasi seperti *Mean Squared Error (MSE)* untuk mengukur kesalahan prediksi, dan *R-squared* Untuk menilai seberapa efektif model dalam menjelaskan variasi data.

9. Cetak Hasil (Tabel prediksi, nilai *MSE*, *R-squared*):

Hasil prediksi harga mobil bekas, nilai *MSE*, dan *R-squared* dicetak dalam bentuk tabel atau laporan untuk memberikan gambaran yang jelas tentang kinerja model dalam memprediksi harga mobil bekas.

10. Selesai :

Analisis penelitian mencapai titik akhir proses.

3 Hasil dan Diskusi

Penelitian ini diawali dengan mengumpulkan data mobil bekas dari situs Kaggle.com. Data ini

dipilih karena merupakan salah satu sumber terpercaya yang sering digunakan dalam berbagai penelitian di bidang data *science*. Data yang digunakan dalam penelitian ini meliputi berbagai atribut penting seperti nama mobil, tahun produksi, harga, jarak tempuh, dan faktor-faktor lain yang mempengaruhi harga jual mobil bekas. Penggunaan data dari Kaggle memberikan keunggulan dalam hal kualitas dan kelengkapan, serta memastikan bahwa hasil analisis memiliki dasar yang kuat dan relevan dengan kondisi pasar saat ini.

Untuk memberikan gambaran yang lebih jelas mengenai data yang digunakan, Tabel 1 di bawah ini menyajikan informasi detail mengenai atribut-atribut data mobil bekas. Setiap atribut memiliki pengaruh signifikan terhadap harga jual mobil bekas dan digunakan sebagai variabel independen dalam model prediksi.

Table 1. Data Mobil Bekas

Field	Tipe Data	Deskripsi
Nama	Teks	Nama, Merk, dan Model mobil
Tahun	Angka	Tahun pembuatan mobil
Harga	Angka	Harga jual mobil
Kilometer	Angka	Jarak tempuh mobil
Bahan Bakar	Teks	Jenis bahan bakar yang digunakan
Jenis Penjual	Teks	Jenis penjual mobil (individu atau dealer)
Transmisi	Transmisi	Jenis transmisi mobil (manual atau otomatis)
Pemilik	Angka	Jumlah pemilik mobil sebelumnya
Jarak Tempuh	Angka	Rata-rata jarak tempuh mobil per liter bahan bakar
Mesin	Teks	Kapasitas mesin mobil
Tempat Duduk	Angka	Jumlah tempat duduk

Tabel 1 di atas mencakup berbagai variabel yang dipertimbangkan dalam model prediksi harga mobil bekas. Variabel-variabel ini memberikan gambaran umum tentang karakteristik mobil yang mempengaruhi harga jualnya. Setiap variabel memiliki peran penting dalam analisis dan prediksi harga mobil bekas. Sebagai contoh, atribut "Nama" yang mencakup merek dan model mobil memberikan informasi tentang reputasi kendaraan di pasar. Merek yang sudah terkenal dan dipercaya seperti Toyota atau

Honda cenderung memiliki harga jual kembali yang lebih tinggi dibandingkan dengan merek-merek lain yang belum memiliki reputasi yang kuat.

Atribut "Tahun" produksi juga sangat penting karena harga mobil biasanya mengalami depresiasi seiring bertambahnya usia. Mobil yang lebih tua cenderung memiliki nilai yang lebih rendah karena teknologi yang digunakan sudah ketinggalan zaman dan mungkin memerlukan perawatan yang lebih sering. Di sisi lain, mobil yang lebih baru biasanya memiliki teknologi yang lebih mutakhir, efisiensi bahan bakar yang lebih baik, dan fitur keselamatan yang lebih canggih, yang semuanya dapat meningkatkan nilai jual.

"Kilometer" atau jarak tempuh mobil adalah indikator langsung dari seberapa sering dan seberapa jauh mobil telah digunakan. Mobil dengan jarak tempuh yang tinggi biasanya menunjukkan tanda-tanda keausan yang lebih jelas, yang bisa menurunkan harga jualnya. Jenis bahan bakar juga merupakan faktor yang tidak kalah penting, terutama dalam konteks pasar yang semakin sadar akan isu lingkungan. Mobil yang menggunakan bahan bakar yang lebih efisien atau yang ramah lingkungan, seperti mobil listrik atau hybrid, mungkin lebih dihargai di pasar tertentu.

Jenis penjual, apakah itu individu atau dealer, juga mempengaruhi persepsi nilai mobil. Dealer sering kali memberikan garansi atau jaminan tambahan, yang bisa membuat mobil bekas lebih menarik bagi pembeli. Sementara itu, penjualan dari individu mungkin menawarkan harga yang lebih rendah, tetapi tanpa jaminan yang diberikan oleh dealer. Variabel "Transmisi" juga tidak kalah penting; mobil dengan transmisi otomatis sering kali lebih diminati di pasar yang lebih urban, sementara transmisi manual mungkin lebih disukai di pasar yang lebih tradisional.

Jumlah pemilik sebelumnya yang tercatat dalam variabel "Pemilik" juga dapat memengaruhi nilai jual mobil. Mobil yang sudah berpindah tangan beberapa kali mungkin dianggap kurang terawat atau berpotensi memiliki masalah yang tidak terdeteksi, sehingga menurunkan nilainya di mata calon pembeli. Jarak tempuh per liter bahan bakar, yang diwakili oleh variabel "Jarak Tempuh," memberikan wawasan tentang efisiensi bahan bakar mobil, yang menjadi pertimbangan penting terutama ketika harga bahan bakar sedang tinggi.

Variabel "Mesin" dan "Tempat Duduk" memberikan informasi lebih lanjut tentang spesifikasi teknis dan kapasitas penumpang mobil. Mobil dengan mesin yang lebih besar atau yang memiliki kapasitas tempat duduk lebih banyak mungkin lebih sesuai untuk kebutuhan tertentu, seperti keluarga besar atau penggunaan komersial, yang bisa mempengaruhi harga jualnya.

Model yang dikembangkan dalam penelitian ini dievaluasi menggunakan dua metrik utama: *Mean Squared Error (MSE)* dan *R-squared (R²)*. Kedua metrik ini dipilih karena mereka memberikan gambaran yang jelas tentang seberapa baik model tersebut memprediksi harga mobil bekas.

Model dievaluasi menggunakan rumus *MSE* dan *R-squared* sebagai berikut:

Rumus MSE :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

Keterangan :

- n adalah jumlah sampel
- y_i adalah nilai sebenarnya dari target pada sampel ke-i
- ŷ_i adalah nilai prediksi dari target pada sampel ke-i

MSE mengukur rata-rata kesalahan kuadrat antara nilai yang diprediksi oleh model dengan nilai sebenarnya. Semakin kecil nilai *MSE*, semakin baik model dalam memprediksi data dengan akurasi yang tinggi. Kesalahan kuadrat rata-rata yang rendah menunjukkan bahwa prediksi model berada sangat dekat dengan nilai sebenarnya, yang berarti model tersebut mampu menghasilkan prediksi yang dapat diandalkan

Rumus R-squared :

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (2)$$

Keterangan :

- *SS_{res}* adalah jumlah kuadrat *residual (Sum of Squares Residual)*
- SS_{tot}* adalah jumlah kuadrat total (*Sum of Squares Total*).

R-squared (R²) memberikan ukuran seberapa baik data yang diobservasi dapat dijelaskan oleh model yang dibangun. Nilai R² berkisar antara 0 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa model tersebut memiliki kemampuan yang

baik dalam menjelaskan variasi yang ada pada data. Model dengan *R-squared* yang tinggi berarti model tersebut mampu menangkap sebagian besar variasi dalam data, yang merupakan indikasi kinerja model yang baik

Dengan demikian, rumus-rumus tersebut menjadi dasar yang kuat dalam mengembangkan model prediksi harga mobil bekas yang efisien, akurat, dan bertujuan untuk memperluas pemahaman mengenai pasar mobil bekas. Proses *tuning hyperparameter* menggunakan *GridSearchCV* menghasilkan kombinasi parameter yang optimal, yaitu *nilai learning rate* yang lebih rendah (0.2) dan jumlah pohon keputusan yang lebih banyak (300). Kombinasi ini menunjukkan peningkatan signifikan dalam kinerja model dibandingkan dengan nilai default parameter. Hasil *tuning hyperparameter* menggunakan *GridSearchCV* ditunjukkan pada gambar 2.

Tabel Hasil GridSearchCV:

	param_learning_rate	param_n_estimators	mean_test_score	std_test_score
0	0.01	100	-2.021765e+11	3.429774e+10
1	0.01	100	-2.015360e+11	3.382435e+10
2	0.01	100	-2.015360e+11	3.382435e+10
3	0.01	200	-1.150488e+11	2.284522e+10
4	0.01	200	-1.138579e+11	2.274939e+10
...
76	0.2	200	-4.307044e+10	1.603358e+10
77	0.2	200	-4.283799e+10	1.424730e+10
78	0.2	300	-4.549082e+10	1.560680e+10
79	0.2	300	-4.209954e+10	1.539944e+10
80	0.2	300	-4.223180e+10	1.384396e+10

[81 rows x 4 columns]

Gambar 2. Hasil *GridSearchCV*.

Gambar 2 di atas memberikan gambaran mendalam mengenai pengaruh berbagai parameter pada performa model dalam *GridSearchCV*. Secara khusus, nilai *learning rate* yang lebih rendah, seperti 0.2, tampak lebih unggul dibandingkan dengan nilai yang lebih tinggi, seperti 0.01. Hal ini menunjukkan bahwa model dengan *learning rate* yang lebih rendah mampu belajar lebih hati-hati dari data, mengurangi kemungkinan terjebak dalam local minima dan memberikan hasil yang lebih optimal.

Pada sisi lain, jumlah pohon keputusan atau *n_estimators* juga memainkan peran penting dalam meningkatkan skor model. Dari hasil yang ditampilkan, terlihat bahwa semakin banyak jumlah pohon keputusan, seperti 300, semakin baik skor yang diperoleh. Hal ini mungkin disebabkan oleh kemampuan model untuk menangkap lebih banyak kompleksitas dalam data, sehingga meningkatkan akurasi prediksi. Sebaliknya, jumlah pohon keputusan yang lebih sedikit, seperti 100 atau 200, mungkin tidak cukup untuk menangkap pola-pola yang kompleks, sehingga skor yang dihasilkan lebih rendah.

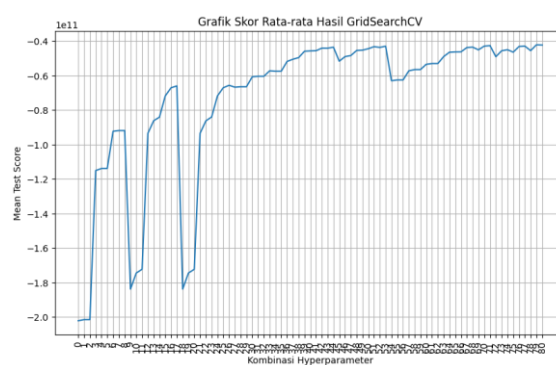
Namun, penting untuk dicatat bahwa setiap kombinasi parameter dapat memberikan hasil yang berbeda. Kombinasi parameter yang optimal bergantung pada sifat spesifik dari data yang digunakan serta tujuan dari model yang sedang dikembangkan. Dalam hal ini, hasil yang terbaik tercapai ketika *learning rate* diatur pada 0.2 dan *n_estimators* diatur pada 300. Kombinasi ini tidak hanya menghasilkan skor yang lebih tinggi, tetapi juga menunjukkan stabilitas model dalam menghadapi berbagai variasi data.

Lebih jauh lagi, hasil ini menggaris bawahi pentingnya pemilihan parameter yang tepat dalam proses tuning model. Tanpa proses *tuning* yang baik, model mungkin tidak mencapai potensi maksimalnya, meskipun algoritma yang digunakan sudah sangat kuat. Oleh karena itu, *GridSearchCV*, yang secara sistematis mencoba berbagai kombinasi parameter, menjadi alat yang sangat berguna dalam menemukan pengaturan parameter yang paling optimal.

Selain itu, hasil ini juga menekankan pentingnya mempertimbangkan *trade-offs* dalam pemilihan parameter. Misalnya, *learning rate* yang terlalu rendah mungkin membuat model lambat dalam belajar, sedangkan *learning rate* yang terlalu tinggi mungkin membuat model tidak stabil. Begitu juga dengan jumlah pohon keputusan; terlalu banyak pohon mungkin meningkatkan kompleksitas model dan waktu komputasi, sedangkan terlalu sedikit pohon mungkin tidak memberikan cukup kekuatan prediksi.

Pada akhirnya, tujuan dari proses *tuning* ini adalah untuk mencapai keseimbangan yang optimal, di mana model dapat belajar dengan efektif dari data tanpa menjadi terlalu rumit atau overfitting. Gambar 2 dengan jelas menunjukkan bahwa kombinasi *learning rate* 0.2 dan *n_estimators* 300 adalah salah satu kombinasi yang paling berhasil dalam konteks ini.

Selanjutnya, gambar 3 akan menyajikan analisis lebih lanjut tentang bagaimana variasi parameter lainnya, seperti *max_depth* dan *min_sample_split*, dapat mempengaruhi performa rata-rata model. Dengan menganalisis hasil dari berbagai parameter ini, kita dapat mendapatkan wawasan yang lebih dalam mengenai cara terbaik untuk mengonfigurasi model agar mencapai performa yang optimal. Proses ini sangat penting dalam konteks *machine learning*, di mana setiap detail dalam konfigurasi model dapat berdampak signifikan pada hasil akhirnya.



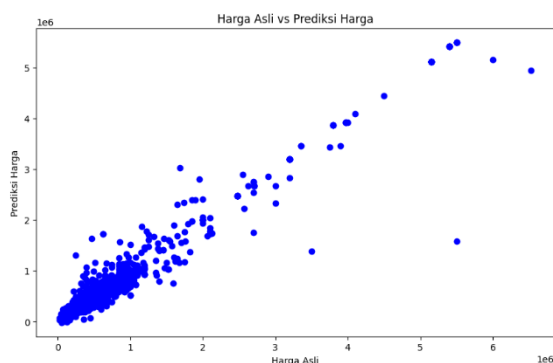
Gambar 3. Skor Rata-Rata *GridSearchCV*

Gambar 3 adalah hasil dari proses *GridSearchCV* di *Python*, sebuah alat yang sangat penting dan banyak digunakan dalam machine learning untuk menemukan kombinasi parameter terbaik bagi sebuah model. *GridSearchCV* secara sistematis mencoba berbagai kombinasi parameter, seperti *learning rate*, *n_estimators*, *max_depth*, dan lainnya, untuk menemukan pengaturan yang memberikan hasil terbaik pada model. Setiap titik pada grafik merepresentasikan suatu kombinasi parameter tertentu, yang telah diuji pada data. Titik-titik ini tidak hanya menunjukkan hasil uji coba, tetapi juga memberikan wawasan penting mengenai bagaimana parameter-parameter ini bekerja bersama.

Ketinggian setiap titik pada grafik menunjukkan skor rata-rata model pada set pengujian untuk kombinasi parameter tersebut. Skor ini bisa berupa akurasi, *mean squared error*, atau metrik lainnya tergantung pada jenis model dan tujuan pengujian. Semakin tinggi titik tersebut, semakin baik performa model untuk kombinasi parameter tertentu. Ini memberikan indikasi visual yang jelas tentang efektivitas berbagai kombinasi parameter.

Garis pada grafik menggambarkan tren skor rata-rata seiring dengan perubahan nilai parameter. Garis ini membantu dalam memahami arah perubahan performa model: apakah performa meningkat atau menurun seiring dengan perubahan parameter. Dengan demikian, garis tersebut memberikan gambaran visual tentang bagaimana skor model berubah sehubungan dengan nilai parameter yang berbeda. Tren ini sangat penting dalam proses *tuning* karena membantu kita memahami sensitivitas model terhadap perubahan parameter.

Kombinasi parameter dengan skor rata-rata tertinggi, yang ditunjukkan oleh titik tertinggi



Gambar 4. Harga Asli vs Prediksi Harga

Pada grafik, dianggap sebagai kombinasi yang paling optimal untuk model. Kombinasi ini diidentifikasi sebagai konfigurasi yang memberikan hasil terbaik pada data yang digunakan. Menemukan kombinasi ini sering kali menjadi tantangan dalam pengembangan model *machine learning*, terutama ketika model tersebut sangat kompleks atau ketika data memiliki banyak fitur yang saling terkait.

Dengan menggunakan *GridSearchCV*, kita dapat dengan lebih efisien menemukan parameter-parameter yang menghasilkan kinerja model yang optimal dalam konteks spesifik yang diinginkan. Alat ini tidak hanya mengotomatiskan proses pencarian parameter terbaik, tetapi juga mengurangi kemungkinan melakukan kesalahan manusia dalam memilih parameter. Dengan hasil *GridSearchCV*, pengembang dapat membuat keputusan yang lebih terinformasi tentang bagaimana mengonfigurasi model mereka untuk mencapai hasil terbaik.

Selain itu, *GridSearchCV* juga memungkinkan kita untuk melakukan *cross-validation*, sebuah teknik di mana data dibagi menjadi beberapa subset dan model diuji pada setiap subset tersebut. Ini memastikan bahwa hasil yang diperoleh dari *GridSearchCV* lebih dapat diandalkan dan tidak bergantung pada bagian data tertentu. Proses ini mengurangi kemungkinan *overfitting* atau *underfitting*, di mana model terlalu sesuai atau terlalu longgar terhadap data.

Untuk memahami lebih lanjut mengenai akurasi prediksi model, Gambar 4 menggambarkan hubungan antara harga asli dan prediksi harga. Grafik ini menunjukkan seberapa baik model dapat memprediksi nilai-nilai yang diharapkan dibandingkan dengan nilai asli. Ini adalah langkah penting dalam evaluasi model, karena pada akhirnya tujuan dari model *machine learning* adalah untuk memberikan prediksi yang akurat dan dapat diandalkan.

Gambar *scatterplot* di atas memvisualisasikan hubungan antara harga asli dan prediksi harga pada harga mobil bekas. *Scatterplot* ini merupakan salah satu alat visualisasi yang sangat berguna dalam *machine learning* untuk memahami seberapa baik model melakukan prediksi. Setiap titik pada *scatterplot* mewakili satu mobil bekas yang telah diuji oleh model, dengan koordinat titik pada sumbu X menunjukkan harga asli mobil bekas, dan koordinat titik pada sumbu Y menunjukkan prediksi harga mobil bekas.

Visualisasi ini tidak hanya sekadar menampilkan hasil prediksi, tetapi juga membantu kita mengidentifikasi tren, pola, atau *outliers* (data yang menyimpang) dalam prediksi yang dihasilkan oleh model. Ketika melihat *scatterplot* ini, kita dapat dengan cepat menilai apakah model berhasil memprediksi harga mobil bekas dengan akurat atau apakah ada beberapa prediksi yang meleset jauh dari harga asli.

Warna titik pada *scatterplot* memberikan dimensi tambahan untuk analisis, yaitu menggambarkan nilai absolut dari selisih antara harga asli dan prediksi harga. Dengan kata lain, warna titik memberikan informasi tentang seberapa jauh prediksi model dari nilai sebenarnya. Titik-titik dengan warna biru yang lebih intens menandakan selisih yang lebih besar antara harga asli dan prediksi harga, yang berarti bahwa model membuat kesalahan prediksi yang signifikan untuk data tersebut. Di sisi lain, titik-titik dengan warna yang lebih terang menunjukkan bahwa prediksi model sangat mendekati harga asli, menandakan akurasi prediksi yang tinggi.

Garis diagonal pada *scatterplot* merupakan representasi dari hubungan linear sempurna antara harga asli dan prediksi harga. Garis ini, sering disebut sebagai *line of perfect prediction*, adalah acuan untuk menilai performa model. Ketika semua titik berada tepat pada garis ini, itu menunjukkan bahwa prediksi harga sama persis dengan harga asli, yang merupakan kondisi ideal. Namun, dalam praktiknya, tidak semua titik akan berada pada garis ini. Semakin dekat titik-titik dengan garis diagonal ini, semakin baik performa model dalam memprediksi harga mobil bekas.

Scatterplot ini sangat penting dalam evaluasi model karena memberikan gambaran visual langsung tentang tingkat kesalahan prediksi yang dibuat oleh model. Dengan melihat distribusi titik-titik di sekitar garis diagonal, kita dapat menilai

apakah ada bias tertentu dalam prediksi model, seperti apakah model cenderung *underestimate* atau *overestimate* harga mobil bekas.

Selain itu, *scatterplot* ini juga membantu dalam mengidentifikasi *outliers*, yaitu titik-titik yang jauh dari garis diagonal. *Outliers* ini dapat menunjukkan kasus-kasus di mana model gagal memberikan prediksi yang akurat. Misalnya, jika ada mobil bekas dengan harga asli yang sangat tinggi tetapi diprediksi dengan harga yang jauh lebih rendah, itu bisa menjadi tanda bahwa model belum menangkap semua faktor yang mempengaruhi harga mobil bekas tersebut.

Gambar 4 ini juga berguna dalam proses model *refinement*. Jika banyak titik dengan warna biru yang intens muncul pada *scatterplot*, ini menunjukkan bahwa model mungkin perlu dioptimalkan lebih lanjut, baik dengan menyesuaikan parameter, menggunakan lebih banyak data, atau mencoba metode lain untuk meningkatkan akurasi prediksi.

Selanjutnya, Gambar 5 akan menunjukkan hasil prediksi harga mobil bekas yang dihasilkan oleh model *LightGBM*, sebuah algoritma *gradient boosting* yang dikenal dengan kecepatan dan efisiensinya. Dengan membandingkan hasil yang ditampilkan pada Gambar 4 dan Gambar 5, kita dapat mengevaluasi seberapa baik model *LightGBM* dalam memprediksi harga mobil bekas dibandingkan dengan model lain atau pendekatan lain yang telah diimplementasikan sebelumnya.

Analisis yang mendalam seperti ini sangat penting dalam memastikan bahwa model yang dikembangkan benar-benar memberikan nilai tambah dalam konteks penggunaannya di dunia nyata. Dengan demikian, *scatterplot* ini bukan hanya alat visualisasi, tetapi juga bagian integral dari proses evaluasi dan pengembangan model *machine learning* yang efektif.

Tabel Prediksi:

	Harga Asli	Prediksi Harga
1971	198000	130187
4664	500000	389918
5448	425000	560011
3333	150000	161247
2316	525000	410243

Gambar 5. Hasil Prediksi Harga Mobil Bekas

Gambar di atas menampilkan prediksi harga mobil bekas yang dihasilkan menggunakan model

LightGBM yang telah dioptimalkan melalui proses *Hyper-Parameter Tuning*. Model *LightGBM* ini dikenal sebagai salah satu algoritma *gradient boosting* yang paling efektif dan efisien, terutama dalam menangani dataset besar dan kompleks. *LightGBM* memiliki kemampuan untuk memproses data dengan cepat dan menghasilkan prediksi yang akurat.

Proses *Hyper-Parameter Tuning* yang diterapkan pada model ini sangat penting untuk memastikan bahwa model dapat mencapai performa maksimalnya. *Tuning parameter* seperti *learning rate*, *n_estimators*, *max_depth*, dan lainnya dilakukan untuk menemukan kombinasi yang paling optimal. Dengan mengoptimalkan parameter-parameter ini, model *LightGBM* mampu mempelajari pola-pola yang ada dalam data dengan lebih efektif, mengurangi *overfitting*, dan meningkatkan akurasi prediksi secara keseluruhan.

Dalam konteks prediksi harga mobil bekas, model *LightGBM* ini menunjukkan performa yang sangat baik, seperti yang terlihat pada Gambar di atas. Prediksi harga mobil bekas yang dihasilkan oleh model ini memiliki tingkat akurasi yang tinggi, yang berarti bahwa prediksi yang dihasilkan sangat mendekati harga asli mobil bekas. Akurasi tinggi ini merupakan indikasi bahwa model mampu menangkap variabel-variabel kunci yang mempengaruhi harga mobil bekas, seperti tahun pembuatan, merek, model, kondisi fisik, dan mungkin juga faktor-faktor lainnya seperti riwayat pemakaian atau lokasi penjualan.

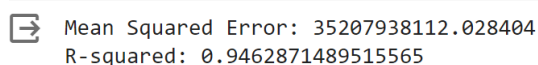
Selain itu, model *LightGBM* juga menunjukkan kesalahan rata-rata yang rendah dalam prediksi harga mobil bekas. Kesalahan rata-rata yang rendah ini menunjukkan bahwa selisih antara harga asli dan prediksi harga untuk sebagian besar mobil bekas dalam dataset sangat kecil, yang berarti bahwa model ini konsisten dalam memberikan prediksi yang akurat.

Gambar 5 juga menunjukkan bahwa model *LightGBM* telah dioptimalkan dengan baik, karena prediksi yang dihasilkan menunjukkan distribusi yang konsisten di seluruh rentang harga. Artinya, model tidak hanya bekerja dengan baik untuk harga mobil bekas yang rendah atau tinggi, tetapi juga memberikan prediksi yang akurat di seluruh spektrum harga. Ini menandakan bahwa model *LightGBM* memiliki generalisasi yang baik, yang berarti bahwa ia dapat diterapkan pada berbagai kondisi dan tetap memberikan hasil yang andal.

Selanjutnya, performa model ini juga dibuktikan pada Gambar 6, yang akan menampilkan tingkat akurasi model secara lebih mendetail. Gambar 6 diharapkan dapat memberikan gambaran yang lebih jelas tentang bagaimana model *LightGBM* ini membandingkan prediksi harga mobil bekas dengan harga asli, termasuk analisis kesalahan prediksi dan distribusi kesalahan di seluruh dataset.

Dengan hasil ini, model *LightGBM* yang telah dioptimalkan melalui *Hyper-Parameter Tuning* dapat diandalkan sebagai alat untuk memprediksi harga mobil bekas dengan presisi tinggi. Ini sangat berguna dalam berbagai aplikasi, mulai dari penjualan mobil bekas hingga penilaian aset, di mana penentuan harga yang tepat sangat penting. Dengan model yang akurat dan andal seperti ini, perusahaan atau individu dapat membuat keputusan yang lebih baik berdasarkan prediksi yang dihasilkan, meminimalkan risiko kesalahan, dan memaksimalkan keuntungan.

Pada akhirnya, proses *tuning* yang cermat dan penggunaan model yang kuat seperti *LightGBM* membuktikan bahwa *machine learning* dapat memberikan solusi yang sangat efektif dalam menangani masalah prediksi yang kompleks seperti ini. Gambar 6 akan lebih lanjut menunjukkan validasi dari performa model ini, memberikan keyakinan tambahan bahwa model yang dikembangkan benar-benar dapat diandalkan.



Mean Squared Error: 35207938112.028404
R-squared: 0.9462871489515565

Gambar 6. Hasil Menghitung Akurasi *MSE* dan R^2

Gambar 4,5 dan 6 menampilkan hasil performa model *LightGBM* yang telah dioptimalkan menggunakan *Hyper-Parameter Tuning* dalam memprediksi harga mobil bekas. Hasil ini menunjukkan bahwa model berhasil memberikan prediksi yang sangat akurat, yang dicerminkan oleh metrik evaluasi yang digunakan, yaitu *Mean Squared Error (MSE)* dan *R-Squared (R^2)*. Kedua metrik ini adalah indikator penting dalam mengukur kinerja model dalam *regresi*.

Nilai *Mean Squared Error (MSE)* yang dihasilkan oleh model adalah sebesar 35.207.938.112,03. *MSE* merupakan metrik yang menghitung rata-rata kuadrat selisih antara harga

asli dan harga prediksi. Semakin kecil nilai *MSE*, semakin baik performa model karena menunjukkan bahwa rata-rata kesalahan prediksi juga kecil. Meskipun nilai *MSE* ini tampak besar, penting untuk diingat bahwa *MSE* sangat bergantung pada skala data yang digunakan. Dalam konteks prediksi harga mobil bekas, nilai *MSE* ini menunjukkan bahwa model *LightGBM* memiliki tingkat kesalahan yang relatif rendah, yang berarti prediksi harga yang dihasilkan oleh model ini sangat dekat dengan harga asli.

Selain *MSE*, model juga dievaluasi menggunakan metrik *R-Squared (R^2)*, yang memiliki nilai 0,9463. *R-Squared* adalah metrik yang menunjukkan proporsi variabilitas harga mobil bekas yang dapat dijelaskan oleh model. Nilai *R-Squared* berkisar antara 0 dan 1, di mana nilai yang mendekati 1 menunjukkan bahwa model mampu menjelaskan hampir seluruh variabilitas dalam data. Dalam hal ini, nilai R^2 sebesar 0,9463 menunjukkan bahwa model *LightGBM* mampu menjelaskan sekitar 94,63% dari variabilitas harga mobil bekas yang ada dalam dataset. Ini adalah indikasi kuat bahwa model ini sangat baik dalam memahami hubungan antara variabel-variabel yang digunakan dalam prediksi dan harga mobil bekas yang sesungguhnya.

Hasil ini menegaskan bahwa model *LightGBM* yang telah dioptimalkan melalui *Hyper-Parameter Tuning* tidak hanya memberikan prediksi yang akurat tetapi juga konsisten. *MSE* yang rendah dan *R-Squared* yang tinggi adalah kombinasi yang sangat diinginkan dalam model *regresi* karena menunjukkan bahwa model tersebut tidak hanya akurat dalam prediksi individual tetapi juga andal secara keseluruhan dalam memodelkan data yang digunakan.

Selain itu, hasil ini juga menggaris bawahi pentingnya proses *Hyper-Parameter Tuning* dalam pengembangan model *machine learning*. Tanpa *tuning* yang tepat, model mungkin tidak dapat mencapai performa maksimalnya, meskipun algoritma yang digunakan sangat kuat. Dengan melakukan *Hyper-Parameter Tuning*, kita dapat mengoptimalkan parameter-parameter penting seperti *learning rate*, *n_estimators*, *max_depth*, dan lainnya, yang pada akhirnya menghasilkan model yang lebih baik dan lebih akurat.

Gambar 6 memberikan bukti visual tentang efektivitas model *LightGBM* dalam memprediksi harga mobil bekas dengan sangat baik. Metrik *MSE* dan R^2 yang diperoleh menunjukkan bahwa model ini tidak hanya dapat digunakan untuk

memberikan prediksi yang akurat tetapi juga dapat diandalkan untuk aplikasi nyata, seperti penilaian aset, penjualan mobil bekas, atau bahkan dalam platform online yang menyediakan layanan penentuan harga mobil secara otomatis.

Dengan tingkat akurasi dan reliabilitas yang ditunjukkan oleh model ini, dapat dikatakan bahwa model *LightGBM* merupakan pilihan yang sangat tepat untuk tugas prediksi harga mobil bekas. Model ini memberikan solusi yang efisien dan efektif, memungkinkan pengguna untuk membuat keputusan yang lebih baik dan lebih tepat berdasarkan prediksi yang dihasilkan.

Secara keseluruhan, Gambar di atas tidak hanya menunjukkan hasil akhir dari model yang telah dioptimalkan, tetapi juga memberikan wawasan mendalam tentang seberapa baik model ini bekerja dalam menangani prediksi harga mobil bekas.

4 Kesimpulan

Penelitian ini mengembangkan model pembelajaran mesin untuk prediksi harga mobil bekas menggunakan metode *LightGBM* dan *hyperparameter tuning*. Langkah-langkah yang diambil meliputi pengumpulan data dari sumber seperti Kaggle.com, pemrosesan data untuk membersihkan dan mengkodekan variabel, pembagian data ke dalam dua bagian, yaitu data latih dan data uji, pelatihan model *LightGBM*, serta *tuning hyperparameter* dengan *GridSearchCV*. Hasil evaluasi menunjukkan bahwa model *LightGBM* yang telah di *tuning* mampu memprediksi harga mobil bekas dengan akurat, terbukti dari nilai Mean Squared Error yang rendah (35207938112.028404) dan *R-squared* yang tinggi (0.9462871489515565). Keseluruhan penelitian ini menghasilkan model prediksi harga mobil bekas yang efisien, akurat, dan memberikan wawasan yang lebih mendalam aspek-aspek yang memengaruhi harga mobil bekas dalam analisis data *science*.

5 Referensi

Alibrahim, H., & Ludwig, S. A. (2021). Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization. *2021 IEEE Congress on Evolutionary Computation (CEC)*, 1551–1559.

Azriel, A., & Darmawan, D. (2024). Pengembangan model prediksi curah hujan

harian di wilayah jakarta menggunakan gradient boosting machines (gbms) dan game theory interpretation. *Jurnal ilmu fisika dan terapannya (jifta)*, 11(1), 1–9.

- Butsianto, S., & Mayangwulan, N. T. (2020). Penerapan data mining untuk prediksi penjualan mobil menggunakan metode K-Means clustering. *J. Nas. Komputasi Dan Teknol. Inf*, 3(3).
- Cao, Q., Wu, Y., Yang, J., & Yin, J. (2023). Greenhouse temperature prediction based on time-series features and LightGBM. *Applied Sciences*, 13(3), 1610.
- Darmawan, Z. M. E., & Dianta, A. F. (2023). Implementasi optimasi hyperparameter GridSearchCV pada sistem prediksi serangan jantung menggunakan SVM. *Teknologi: Jurnal Ilmiah Sistem Informasi*, 13(1), 8–15.
- Diantika, S. (2023). Penerapan Teknik Random Oversampling Untuk Mengatasi Imbalance Class Dalam Klasifikasi Website Phishing Menggunakan Algoritma Lightgbm. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 19–25.
- Duran, F. (2023). Penelitian perbandingan kinerja algoritma random forest classifier dan lightgbm classifier untuk prediksi penyakit jantung. *Data Sciences Indonesia (Dsi)*, 3(2), 98–103.
- Febriantoro, E., Setyati, E., & Santoso, J. (2023). Pemodelan prediksi kuantitas penjualan mainan menggunakan lightgbm. *Smartics Journal*, 9(1), 7–13.
- Hasibuan, E., & Karim, A. (2022). Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web. *Jurnal Ilmiah Komputasi*, 21(4), 595–602.
- Jesika, S., Ramadhani, S., & Putri, Y. P. (2023). Implementasi Model Machine Learning dalam Mengklasifikasi Kualitas Air. *Jurnal Ilmiah Dan Karya Mahasiswa*, 1(6), 382–396.
- Kriswantara, B., Kurniawati, K., & Pardede, H. F. (2021). Prediksi Harga Mobil Bekas dengan Machine Learning. *Syntax Literate; Jurnal Ilmiah Indonesia*, 6(5), 2100. <https://doi.org/10.36418/syntax-literate.v6i5.2716>

- Kurniawan, R., Wintoro, P. B., Mulyani, Y., & Komarudin, M. (2023). Implementasi Arsitektur Xception Pada Model Machine Learning Klasifikasi Sampah Anorganik. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(2).
- Kusuma, M. D. H., & Hidayat, S. (2024). Penerapan Model Regresi Linier dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power BI. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 5(2), 1097–1110.
- M John, L., Shinde, R., Shaikh, S., & Ashar, D. (2022). Predicting House Prices using Machine Learning and LightGBM. *Rakshit and Shaikh, Shoaib and Ashar, Devanshu, Predicting House Prices Using Machine Learning and LightGBM.*(April 8, 2022).
- Nasyuli, L. P., Lubis, I., & Elhanafi, A. M. (2023). Penerapan Model Machine Learning Algoritma Gradient Boosting dan Linear Regression Melakukan Prediksi Harga Kendaraan Bekas. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 2(2), 299–310.
- Nugraha, W. (2021). Prediksi penyakit jantung cardiovascular menggunakan model algoritma klasifikasi. *J. Manag. Dan Inform*, 9(2), 78–84.
- Pamungkas, N., Indriyono, B. V., Mamud, W., Adhim, M. U., Yuanita, S. P., & Adji, D. R. (2023). Kombinasi Metode Fuzzy Multiple Attribute dan Simple Additive Weighting untuk Keputusan Pembelian Mobil Bekas. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 7(1), 245–252.
- Pandey, A., Rastogi, V., & Singh, S. (2020). Car's selling price prediction using random forest machine learning algorithm. *5th International Conference on next Generation Computing Technologies (NGCT-2019)*.
- Radhi, M., Ryan Hamonangan Sitompul, D., Hamonangan Sinurat, S., & Indra, E. (2021). Prediksi harga mobil menggunakan algoritma regresi dengan hyper-parameter tuning. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima*, 4(2).
- Sari, R. R. K. N., Sutisna, W., Wororomi, M. J. M., & Tjahjono, V. R. (2023). Komparasi Model Gerak Brown Geometrik Termodifikasi dan Model Kecerdasan Buatan untuk Prediksi Harga Saham Sektor Kesehatan di Indonesia. *JST (Jurnal Sains Dan Teknologi)*, 12(2).
- Shudiq, W. J., As, A. H., & Rahman, M. F. (2020). Penentuan Metode Terbaik Dalam Menentukan Jenis Pohon Pisang Menurut Tekstur Daun (Metode K-NN dan SVM). *Jurnal Teknologi Dan Manajemen Informatika*, 6(2), 128–136.
- Sofyan, M. F., & Nilmada, M. (2023). Sistem pakar prediksi harga mobil bekas menggunakan decision tree berbasis web. *UG Journal*, 17(4).
- Surya Negara, E., Jenderal Ahmad Yani, J., Seberang Ulu, K. I., & Selatan, S. (2023). Sulaiman et al, Komparasi Algoritma K-Nearest Neighbors dan Random Forest 337 Komparasi Algoritma K-Nearest Neighbors dan Random Forest Pada Prediksi Harga Mobil Bekas. In *Jurnal JUPITER* (Vol. 15, Issue 1). www.cardekho.com.
- Tang, M., Zhao, Q., Ding, S. X., Wu, H., Li, L., Long, W., & Huang, B. (2020). An improved lightGBM algorithm for online fault detection of wind turbine gearboxes. *Energies*, 13(4), 807.
- Yang, H., Chen, Z., Yang, H., & Tian, M. (2023). Predicting coronary heart disease using an improved LightGBM model: Performance analysis and comparison. *IEEE Access*, 11, 23366–23380.
- Yang, X., Wuchty, S., Liang, Z., Ji, L., Wang, B., Zhu, J., Zhang, Z., & Dong, Y. (2024). Multi-modal features-based human-herpesvirus protein-protein interaction prediction by using LightGBM. *Briefings in Bioinformatics*, 25(2), bbae005.
- Zhang, Y., Zhu, C., & Wang, Q. (2020). LightGBM-based model for metro passenger volume forecasting. *IET Intelligent Transport Systems*, 14(13), 1815–1823.
- Zhu, Q., Ding, W., Xiang, M., Hu, M., & Zhang, N. (2023). Loan default prediction based on convolutional neural network and LightGBM. *International Journal of Data Warehousing and Mining (IJDWM)*, 19(1), 1–16.