



Perbandingan Metode Random Forest dan LightGBM untuk Prediksi Harga Berlian: Pendekatan Probabilistik dan Statistik

Comparison Of Random Forest And LightGBM Methods For Diamond Price Prediction: A Probabilistic And Statistical Approach

Daud Aldo Santoso¹, Yosefina Finsensia Riti^{1*}

¹ Program Studi Informatika, Fakultas Teknik, Universitas Katolik Darma Cendika, Surabaya, Indonesia

Abstrak: Harga berlian di pasar memang sangat gampang berubah-ubah dan sangat terpengaruh oleh kondisi fisiknya. Hal ini bikin urusan menebak harganya jadi tantangan tersendiri buat banyak orang, apalagi kalau cuma pakai tebakan biasa. Masalah utamanya adalah orang-orang sering kesulitan menaksir harga yang pas secara manual, dan ini bisa bikin rugi besar kalau tebakannya meleset jauh. Nah, lewat studi ini, peneliti punya tujuan pengen melihat sekaligus mengadu langsung kemampuan dua program komputer pintar, yakni algoritma Machine Learning Random Forest dan LightGBM, saat dipakai buat meramal harga biar lebih akurat dan nggak asal tebak. Untuk cara kerjanya, ada sekitar 53.940 riwayat penjualan berlian yang digunakan sebagai bahan belajar. Data ini sudah lengkap dengan sembilan ciri utamanya, semisal berat karat, kualitas potongan, pigmen warna, sampai ukuran fisiknya. Sebelum mulai dihitung oleh komputer, baris data yang isinya aneh atau nggak masuk akal langsung dihapus bersih. Habis itu, keterangan yang masih berupa tulisan diganti ke dalam wujud angka biar gampang diproses. Selanjutnya, keseluruhan data ini dipecah dengan porsi 80 persen untuk bahan belajar si program, dan 20 persen sisanya disimpan murni cuma buat alat tes seberapa pintar tebakan mereka. Buat ngukur ketepatan tebakan, peneliti pakai tiga patokan hitungan, yaitu R-squared (R^2), Mean Absolute Error (MAE), dan Root Mean Squared Error (RMSE). Dari hasil tesnya, Random Forest ternyata jauh lebih jago dengan skor akurasi R^2 yang tembus angka 0.9835, serta rata-rata tebakan meleset (MAE) yang sangat kecil, cuma di kisaran \$38.89 dan RMSE \$77.24. Kemampuan ini sukses telak mengalahkan program LightGBM yang cuma dapat skor R^2 0.9830, tebakan meleset (MAE) \$41.28, serta RMSE \$78.27. Intinya, hasil penelitian ini membuktikan dengan jelas kalau sistem kerja bareng-bareng atau ensemble milik Random Forest jauh lebih awet, stabil, dan paling akurat buat urusan nebak harga berlian di pasaran.

Keywords: Algoritma Regresi; Harga Berlian; LightGBM; Machine Learning; Random Forest.

Abstract: Figuring out diamond prices in the market is actually pretty tricky because the prices go up and down a lot and depend heavily on what the diamond looks like physically. The main problem here is that people often struggle to guess the right price manually, which can easily lead to losing money if they guess wrong. Through this study, we really want to test and directly compare two smart Machine Learning tools, which are the Random Forest and LightGBM algorithms, to see which one is actually better and more reliable at forecasting these prices without just blindly guessing. For how we did it, we used a huge dataset containing about 53,940 past diamond sales as study material. This data came complete with nine main features like carat weight, cut quality, color pigment, and physical size. Before doing the math with the computer, we quickly threw away any weird or impossible data rows to keep it clean. Then, we changed all the text descriptions into numbers so the program could read them easily. Next, we split the whole dataset up so that 80 percent was used to teach the program, and the remaining 20 percent was kept purely for testing how smart their guesses really were. To check how accurate their guesses were, we used three main measurements: R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The test results clearly showed that Random Forest did a much better job overall. It hit a super high R^2 accuracy score of 0.9835, and its guesses were only off by an average of \$38.89 (MAE) and an RMSE of \$77.24. This easily beat the LightGBM program, which only got an R^2 score of 0.9830, an MAE of \$41.28, and an RMSE of \$78.27. In short, this research proves that the team-work or ensemble system in Random Forest is way more reliable, stable, and accurate for predicting diamond prices in the market.

Keywords: Regression Algorithm; Diamond Prices; LightGBM; Machine Learning; Random Forest.

* Corresponding author : yosefina.riti@ukdc.ac.id

<https://doi.org/10.51132/teknologika.v16i1>

Received : 30-04-2026

Accepted : 20-05-2026

Available online : 31-05-2025

1. Pendahuluan

Berlian adalah batu mulia yang dikenal luas karena keindahan dan nilainya dalam berbagai sektor, termasuk perhiasan dan investasi. Namun, penentuan harga berlian cukup rumit karena dipengaruhi oleh banyak faktor yang tidak selalu berbanding lurus, seperti karat (*carat*), potongan (*cut*), warna (*color*), dan tingkat kejernihan (*clarity*) [1]. Karena kerumitan ini, memiliki metode prediksi harga yang akurat sangat diperlukan [2]. Prediksi yang baik berguna untuk membantu pembeli maupun penjual dalam memperkirakan harga yang wajar dan membuat keputusan transaksi yang lebih tepat [3].

Dalam kajian literatur terkait, implementasi algoritma *machine learning* telah banyak digunakan untuk memecahkan masalah prediksi harga berlian ini. Berbagai penelitian sebelumnya telah menguji bermacam metode, mulai dari algoritma dasar seperti Regresi Linier [4], algoritma berbasis pohon seperti *Decision Tree* [3], penerapan model gabungan tingkat lanjut (*enhanced ensemble learning*), hingga algoritma yang lebih kompleks seperti *Support Vector Machine* dan *Neural Networks* [1]. Tinjauan terhadap berbagai penelitian ini menunjukkan bahwa pendekatan komputasi sangat krusial dalam menentukan nilai aset berharga seperti komoditas berlian dan emas.

Pemilihan algoritma yang tepat sangatlah penting. Sebelumnya, algoritma LightGBM telah terbukti menjadi salah satu metode yang andal dalam memprediksi harga pada sektor data lain, seperti estimasi harga mobil bekas [5]. Selain itu, perbandingan beberapa metode regresi juga lumrah dilakukan untuk mencari hasil prediksi yang paling optimal, contohnya penggunaan metode ARIMA dan *Least Square* untuk peramalan harga aset [6].

Mengikuti pendekatan tersebut, penelitian ini akan membandingkan kinerja LightGBM dengan algoritma yang menduduki peringkat terbaik khusus di sektor data berlian. Berdasarkan tinjauan dari berbagai pengujian algoritma sebelumnya, algoritma Random Forest secara konsisten diakui menunjukkan hasil yang paling superior [4]. Random Forest terbukti mampu menghasilkan tingkat akurasi (*R-squared*) tertinggi, sekaligus menekan tingkat kesalahan rata-rata (MAE dan RMSE) hingga titik terendah dibandingkan algoritma *machine learning* lainnya [3], [1].

Oleh karena itu, penelitian ini menjadikan Random Forest sebagai acuan utama untuk diadu secara komparatif dengan LightGBM. Berdasarkan rumusan masalah tersebut, tujuan utama dari riset ini adalah melihat perbandingan performa antara algoritma Random Forest dan LightGBM saat dipakai untuk menebak harga berlian, tentunya dengan mengandalkan hitungan statistik dan peluang. Evaluasi difokuskan pada kemampuan kedua model dalam mengenali pola data historis serta kemampuannya dalam memperkecil persentase kesalahan tebakan harga. Melalui pengujian ini, peneliti berharap bisa memberikan sudut pandang yang luas tentang mengenai arsitektur algoritma regresi mana yang paling akurat, stabil, dan efisien untuk diterapkan dalam sistem prediksi harga pasar komoditas berlian.

2. Metodologi

2.1 Konsep Regresi dan Prediksi Harga

Secara konsep probabilistik maupun statistik, regresi merupakan teknik yang diaplikasikan untuk mengonstruksi model korelasi antara sebuah variabel target (seperti nominal harga) terhadap berbagai variabel independen (faktor-faktor pembentuk, seperti spesifikasi karat, proporsi dimensi, serta kualitas) [7]. Berbeda dengan klasifikasi yang memprediksi label diskrit, model regresi bertujuan untuk menghasilkan prediksi berupa nilai angka yang berkelanjutan (*continuous values*). Dalam konteks pasar komoditas, metode regresi statistik difokuskan untuk menemukan fungsi pemetaan $f(X)$ yang dapat meminimalkan selisih antara harga prediksi dan harga pasar aktual [6].

2.2 Algoritma Random Forest

Cara kerja algoritma Random Forest adalah dengan menggabungkan banyak *decision tree*

sekaligus. Tujuannya untuk memproduksi tingkat tebakan yang presisi sekaligus meminimalisir fluktuasi hasil [8]. Konsep operasional utamanya mengandalkan teknik *Bagging (Bootstrap Aggregating)*. Pada fase pembelajaran, model ini secara otomatis menciptakan beragam variasi sampel data acak dari himpunan data pokok (melalui metode pengembalian), untuk kemudian melatih masing-masing pohon keputusan menggunakan setiap variasi sampel tersebut [9].

Khusus untuk penyelesaian masalah regresi, tebakan akhir diformulasikan dengan cara mengkalkulasi nilai tengah (*mean*) dari seluruh prediksi yang dihasilkan oleh kumpulan pohon di dalam sistem [10]. Secara matematis, prediksi Random Forest untuk sebuah data uji x direpresentasikan melalui persamaan berikut:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B f_b(x) \quad (1)$$

Keterangan:

- \hat{y} = Nilai harga prediksi akhir.
- B = Jumlah total pohon keputusan (*decision trees*) yang dibentuk.
- $f_b(x)$ = Hasil prediksi dari pohon keputusan ke- b .

Berdasarkan persamaan (1) kelebihan utama dari agregasi rata-rata ini adalah kemampuannya dalam menurunkan varians model secara drastis tanpa meningkatkan bias. Hal ini membuat Random Forest sangat tangguh terhadap data pencilan (*outliers*) dan secara efektif mencegah terjadinya *overfitting* [11].

2.3 Algoritma LightGBM

LightGBM, sebuah metode komputasi canggih yang mengadopsi prinsip *gradient boosting* pada struktur pohon keputusan [12]. Pendekatan pembentukan pohon pada model ini bersifat sekuensial, yang menjadikannya sangat berbeda dari mekanisme independen milik Random Forest. Artinya, setiap pohon yang baru diciptakan memiliki misi spesifik untuk memperbaiki sisa kesalahan (*residual error*) yang ditinggalkan oleh generasi pohon sebelumnya [13]. Strategi utama yang digunakan adalah *leaf-wise growth* (memprioritaskan perluasan daun dengan tingkat kerugian tertinggi), yang diklaim jauh lebih menghemat waktu pemrosesan dan penggunaan memori ketika berhadapan dengan data berskala masif [12].

Model prediksi pada *gradient boosting* dibentuk secara iteratif. Jika $F_M(x)$ adalah prediksi harga akhir pada iterasi ke- M , maka persamaan pembentukan modelnya adalah [13]:

$$F_M(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (2)$$

Keterangan:

- $F_M(x)$ = Model prediksi kuat pada iterasi akhir.
- $F_{m-1}(x)$ = Model prediksi dari iterasi sebelumnya.
- $h_m(x)$ = Pohon keputusan dasar (*weak learner*) baru yang ditambahkan.
- γ_m = Nilai *learning rate* atau bobot langkah untuk meminimalkan *loss function*.

Pada persamaan (2) keunggulan matematis ini memungkinkan LightGBM untuk menangkap pola regresi yang sangat kompleks dengan penggunaan memori yang minimal [14].

2.4 Metrik Evaluasi Statistik

Untuk mengukur seberapa baik kinerja pendekatan probabilistik dari model regresi dalam memprediksi harga, penelitian ini menggunakan tiga metrik evaluasi utama secara komparatif [7]:

1. R-squared (R^2)

Koefisien determinasi, yang lazim disebut R^2 , berfungsi sebagai indikator matematis untuk memperlihatkan seberapa besar pengaruh variabel prediktor (seperti karakteristik fisik berlian) dalam menjelaskan perubahan pada variabel target (harga jual) [7]. Rumusnya adalah:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (3)$$

Berdasarkan persamaan (3) y_i adalah harga aktual, \hat{y}_i adalah harga prediksi, dan \bar{y} adalah rata-rata harga aktual. Nilai R^2 mendekati 1 menandakan akurasi model yang sangat tinggi.

2. Mean Absolute Error (MAE)

Secara konseptual, MAE adalah alat ukur untuk menghitung nilai tengah dari penyimpangan mutlak antara angka hasil prediksi komputer dengan angka kenyataan di lapangan [7]. Rumus MAE adalah seperti persamaan (4):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

3. Root Mean Squared Error (RMSE)

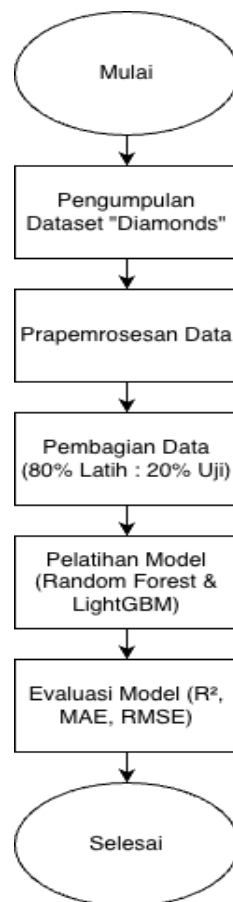
RMSE beroperasi dengan cara mengakarkan nilai rata-rata dari kuadrat selisih antara tebakan model dan harga asli. Proses pengkuadratan ini secara otomatis akan memberikan bobot penalti yang jauh lebih berat pada hasil prediksi yang melenceng terlalu ekstrem [7]. Rumus RMSE dapat dilihat di persamaan (5):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

Model prediksi harga yang optimal adalah model yang memiliki *R-squared* tertinggi, dengan nilai MAE dan RMSE yang paling kecil.

2.5 Alur Penelitian

Secara umum, tahapan penelitian ini disusun secara terstruktur, diawali dengan proses ekstraksi data dan diakhiri dengan tahap pengujian performa. Desain metodologi ini diterapkan secara ketat guna menjamin objektivitas dan validitas saat membandingkan kapabilitas algoritma Random Forest dengan LightGBM.



Gambar 1. Alur Metodologi Penelitian

2.6 Pengumpulan dan Prapemrosesan Data

Penelitian ini menggunakan *dataset* publik sekunder bernama "Diamonds" yang diakses melalui repositori Kaggle [15]. *Dataset* ini memiliki total hampir 54.000 baris data rekam jejak penjualan berlian dengan 10 fitur utama. Fitur target (*dependent variable*) yang akan diprediksi adalah harga (*price*) dalam mata uang dolar AS. Sementara itu, terdapat 9 fitur prediktor (*independent variables*) yang merepresentasikan spesifikasi fisik berlian, antara lain:

1. **Carat:** Berat fisik berlian.
2. **Cut:** Kualitas potongan berlian (Fair, Good, Very Good, Premium, Ideal).
3. **Color:** Tingkat warna berlian, dari J (paling kuning) hingga D (paling jernih/putih).
4. **Clarity:** Tingkat kejernihan dari cacat internal (I1 hingga IF).
5. **X, Y, Z:** Dimensi panjang, lebar, dan kedalaman berlian dalam milimeter.
6. **Depth & Table:** Persentase proporsi dimensi berlian.

Pada tahap prapemrosesan, data mentah ini dibersihkan dari anomali, seperti menghapus baris data yang memiliki nilai dimensi (X, Y, atau Z) sebesar 0, karena tidak memiliki nilai logis di dunia nyata. Selanjutnya, fitur kategorikal yang berupa teks (*cut*, *color*, dan *clarity*) diubah menjadi angka numerik menggunakan teknik *Ordinal Encoding*. Teknik ini dipilih karena ketiga fitur tersebut memiliki tingkatan hierarki kualitas yang pasti (misalnya, *Ideal* lebih baik dari *Fair*).

2.7 Pembentukan dan Pelatihan Model

Data yang telah dibersihkan secara matematis dipecah menjadi dua bagian dengan perbandingan 80 banding 20. Pelatihan model menggunakan 80% dari total semua data, sedangkan sisa 20% nya

disimpan khusus untuk tahap pengujian (*data testing*). Tahap pelatihan dilakukan menggunakan bahasa pemrograman Python. Model Random Forest dibangun dengan memanfaatkan *library* Scikit-Learn, sedangkan model LightGBM dibangun menggunakan *library* resmi LightGBM. Kedua algoritma ini "belajar" memetakan pola hubungan antara spesifikasi fisik berlian (seperti berat karat dan kualitas potongan) terhadap harga jualnya menggunakan 80% data pelatihan tersebut.

2.8 Evaluasi dan Perbandingan Model

Untuk mengukur efektivitas prediksi setelah proses pelatihan selesai, kedua algoritma dihadapkan pada 20% *data testing* sebagai instrumen evaluasi komparatif. Tiga tolok ukur statistik yang diimplementasikan untuk membandingkan Random Forest dan LightGBM meliputi *R-squared* (R^2), *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE). Kesimpulan mengenai model terbaik dalam studi ini akan dijabarkan pada algoritma yang sukses memperoleh nilai kesalahan minimum pada RMSE dan MAE, serta persentase R^2 yang paling mendekati angka sempurna.

3. Hasil dan Pembahasan

3.1 Hasil Pengumpulan dan Prapemrosesan Data

Berdasarkan *dataset* "Diamonds", diperoleh data mentah yang berisi rekam jejak harga dan spesifikasi fisik berlian. Pada tahap prapemrosesan, dilakukan pembersihan data secara statistik untuk membuang anomali atau *outlier*, khususnya pada baris data yang memiliki dimensi fisik nol (0 mm) karena tidak logis di dunia nyata. Setelah pembersihan, fitur kategorikal yang berupa teks seperti *cut* (kualitas potongan), *color* (warna), dan *clarity* (tingkat kejernihan) berhasil diubah menjadi format angka numerik (*ordinal encoding*) sehingga seluruh matriks data siap diproses oleh algoritma regresi.

Sebagai contoh perubahan data, Tabel 1 menampilkan perbandingan lima baris pertama dari *dataset* sebelum dan sesudah diproses. Kolom data yang awalnya berisi teks, seperti *Cut* (Ideal, Premium, Good) dan *Color* (E, I, J), seluruhnya sudah diubah menjadi angka.

Tabel 1. Contoh Data Sebelum dan Sesudah Prapemrosesan

Status Data	Carat	Cut (Potongan)	Color (Warna)	Clarity	Price (\$)
Sebelum (Mentah)	0.23	Ideal	E	SI2	326
Sesudah (Proses)	0.23	4	1	3	326
Sebelum (Mentah)	0.21	Premium	E	SI1	326
Sesudah (Proses)	0.21	3	1	2	326
Sebelum (Mentah)	0.29	Premium	I	VS2	334
Sesudah (Proses)	0.29	3	5	4	334

3.2 Hasil Pembentukan dan Pelatihan Model

Data yang telah dibersihkan kemudian dipisah menjadi 2 kelompok. 80% dari total data sebagai data pelatihan untuk mempelajari pola pergerakan harga. Sisa 20% data digunakan sebagai data pengujian (*testing data*). Pada tahap pelatihan, kedua model secara sukses membentuk arsitektur pohon keputusannya masing-masing pada lingkungan pemrograman Python. Setelah dilatih, kedua model digunakan untuk menebak harga pada data pengujian, dan hasilnya dievaluasi secara statistik.

3.3 Evaluasi dan Perbandingan Model

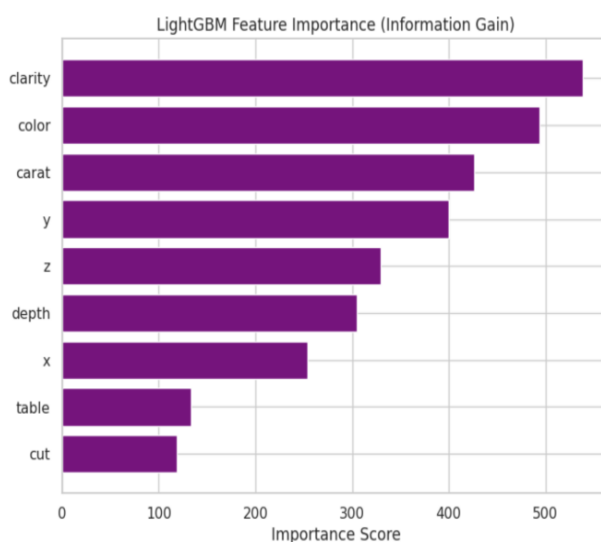
Hasil pengujian menunjukkan kinerja prediksi yang sangat baik dari kedua algoritma. Namun, evaluasi komparatif mencatat adanya perbedaan tingkat keakuratan dan stabilitas antara Random Forest dan LightGBM. Perbandingan hasil metrik evaluasi dari kedua model dirangkum pada Tabel 2.

Tabel 2. Perbandingan Kinerja Prediksi Model

Algoritma	R-squared	MAE	RMSE
LightGBM	0.9830	\$41.28	\$78.27
Random Forest	0.9835	\$38.89	\$77.24

Jika melihat perbandingan angka pada Tabel 2, algoritma Random Forest tampil sebagai model yang lebih baik. Hal ini terlihat dari perolehan nilai akurasi R-squared tertinggi (0.9835) dengan tingkat kesalahan (MAE dan RMSE) yang paling minimal. Keunggulan ini membuktikan bahwa pendekatan *ensemble bagging* lebih stabil dalam menangani data fluktuatif dibandingkan LightGBM.

Selain membandingkan tingkat kesalahan, model juga dievaluasi untuk melihat fitur mana yang paling mempengaruhi pembentukan harga. Hubungan tingkat pengaruh antar fitur tersebut divisualisasikan pada Gambar 2.



Gambar 2. Grafik Tingkat Kepentingan Fitur terhadap Harga

Berdasarkan bukti grafik *Information Gain* pada Gambar 2, terungkap pola yang menarik di mana fitur tingkat kejernihan (*Clarity*) dan warna (*Color*) justru memberikan sumbangan informasi paling tinggi dalam menentukan harga berlian, mengalahkan faktor berat fisik. Fitur *Carat* berada di peringkat ketiga, disusul oleh dimensi fisik lebar (*Y*) dan kedalaman (*Z*). Hal ini membuktikan bahwa dalam membentuk pohon keputusannya, algoritma sangat mengandalkan kombinasi kualitas visual berlian (kejernihan dan warna) sebagai pembeda utama harga, bukan sekadar ukuran fisiknya saja.

4. Kesimpulan

Kesimpulan utama dari riset ini menunjukkan keunggulan performa algoritma Random Forest dibandingkan LightGBM dalam konteks estimasi harga berlian. Model Random Forest mencapai

akurasi R^2 sebesar 0.9835 dengan tingkat kesalahan absolut (MAE) sebesar \$38.89, mengalahkan capaian LightGBM yang berada pada angka R^2 0.9830 dan MAE \$41.28. Hasil tersebut menegaskan bahwa pendekatan *ensemble bagging* pada Random Forest memiliki stabilitas yang lebih tinggi dalam memproses karakteristik fisik berlian untuk menghasilkan prediksi harga pasar yang akurat.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi Ilmu Informatika, Fakultas Teknik, Universitas Katolik Darma Cendika atas dukungan fasilitas yang diberikan selama proses penyusunan penelitian ini. Penelitian ini tidak didanai oleh lembaga / pihak manapun.

Daftar Pustaka

- [1] W. Alsuraihi, E. Al-hazmi, K. Bawazeer, dan H. Alghamdi, "Machine Learning Algorithms for Diamond Price Prediction," dalam *Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing*, dalam IVSP '20. New York, NY, USA: Association for Computing Machinery, Mei 2020, hlm. 150–154. doi: 10.1145/3388818.3393715.
- [2] H. Ranglani, "<p>Integrating Machine Learning for Diamond Price Prediction and Distinguishing Natural Diamonds from Lab Grown: A Unified Approach</p>," 1 Maret 2025, *Social Science Research Network, Rochester, NY*: 5541978. doi: 10.2139/ssrn.5541978.
- [3] A. Mankawade, C. Kokate, K. Soman, A. Mohite, A. Vispute, dan O. More, "Diamond Price Prediction Using Machine Learning Algorithms," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, hlm. 4867–4871, Mei 2023, doi: 10.22214/ijraset.2023.52741.
- [4] A. Agustina, Y. N. Putri, H. Nisah, S. D. Raihanah, A. Najib, dan V. Tundjungsari, "Analisis Prediksi Harga Berlian Menggunakan Metode Regresi Linier," *JATI J. Mhs. Tek. Inform.*, vol. 10, no. 1, hlm. 1825–1832, Feb 2026, doi: 10.36040/jati.v10i1.17567.
- [5] M. A. Aulady, A. H. As, dan Z. Arifin, "Improve Metode Lightgbm untuk Prediksi Harga Mobil Bekas Menggunakan Hyper-Parameter Tuning," *TRILOGI J. Ilmu Teknol. Kesehat. Dan Hum.*, vol. 5, no. 3, hlm. 456–467, Sep 2024, doi: 10.33650/trilogi.v5i3.9000.
- [6] D. Anggelia, Y. F. Riti, dan P. W. Siswanto, "Analisis Perbandingan Metode Arima Dan Least Square Untuk Prediksi Harga Emas: Pendekatan Probabilistik Dan Statistik," *J. Sist. Inf. Dan Inform. Simika*, vol. 7, no. 1, hlm. 95–103, Mar 2024, doi: 10.47080/simika.v7i1.3197.
- [7] T. Hastie, R. Tibshirani, dan J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. dalam Springer series in statistics. New York, NY: Springer, 2009.
- [8] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, hlm. 5–32, Okt 2001, doi: 10.1023/A:1010933404324.
- [9] E. S. Lestari dan I. Astuti, "Penerapan Random Forest Regression Untuk Memprediksi Harga Jual Rumah Dan Cosine Similarity Untuk Rekomendasi Rumah Pada Provinsi Jawa Barat," *J. Ilm. FIFO*, vol. 14, no. 2, hlm. 131, Nov 2022, doi: 10.22441/fifo.2022.v14i2.003.
- [10] I. M. G. A. B. Putra dan I. K. G. Suhartana, "Implementasi Algoritma Random Forest Regression dalam Sistem Prediksi Harga Rumah di Jabodetabek," *J. Nas. Teknol. Inf. Dan Apl.*, vol. 4, no. 1, hlm. 27–38, Nov 2025, doi: 10.24843/JNATIA.2025.v04.i01.p04.
- [11] P. S. Saputra, S.Ti., M.Kom. dan I. P. G. A. Sudiatmika, "Analisis Prediksi Harga Smartphone Tahun 2023 Menggunakan Model Random Forest Regression Berdasarkan Fitur-Fitur Spesifikasi Teknis," *KOMTEKS*, vol. 3, no. 2, hlm. 13–17, Jan 2025, doi: 10.37637/komteks.v3i2.2233.
- [12] G. Ke dkk., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," dalam *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Diakses: 30 Maret 2026. [Daring]. Tersedia pada: https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

- [13] J. H. Friedman, “Greedy function approximation: A gradient boosting machine.,” *Ann. Stat.*, vol. 29, no. 5, Okt 2001, doi: 10.1214/aos/1013203451.
- [14] E. Febriantoro, E. Setyati, dan J. Santoso, “Pemodelan Prediksi Kuantitas Penjualan Mainan Menggunakan LightGBM,” *SMARTICS J.*, vol. 9, no. 1, hlm. 7–13, Apr 2023, doi: 10.21067/smartics.v9i1.8279.
- [15] “Diamonds.” Diakses: 30 Maret 2026. [Daring]. Tersedia pada: <https://www.kaggle.com/datasets/shivam2503/diamonds>