

Analisis Prediksi Performa Produk Teh Menggunakan Algoritma *Regresi Linear* dan *Random Forest Regressor*

¹James Januard Lada', ²Dwi Nurani, ³Alfie Nur Rahmi, ⁴Supriatin
^{1,2,3,4}Universitas Amikom Yogyakarta
Yogyakarta, Indonesia

¹jamesjanuard@students.amikom.ac.id,
²dwinurani@amikom.ac.id,³alfienurrahmi@amikom.ac.id, ⁴supriatin@amikom.ac.id

*Penulis Korespondensi

Diajukan : 29/12/2025
Diterima : 06/01/2026
Dipublikasi : 09/01/2026

ABSTRAK

Penelitian ini berfokus pada analisis prediksi performa produk teh menggunakan algoritma regresi linear dan Random Forest Regressor (RFR). Masalah yang diangkat adalah ketidakpastian dalam memprediksi performa penjualan produk teh yang dapat berdampak pada strategi pemasaran dan pengambilan keputusan di industri teh. Ketidakakuratan dalam proyeksi penjualan dapat menyebabkan kerugian finansial dan menurunkan daya saing produk di pasar. Oleh karena itu, penting untuk menerapkan metode analisis yang tepat untuk meningkatkan akurasi prediksi. Metode penelitian yang digunakan meliputi pengumpulan data dari platform Kaggle, di mana variabel yang dianalisis mencakup harga produk, rating rata-rata, dan jumlah ulasan. Penelitian ini menerapkan algoritma regresi linear untuk memahami hubungan linear antar variabel, serta RFR untuk menangani data dengan kompleksitas tinggi. Proses analisis mengikuti tahapan Knowledge Discovery in Database (KDD) yang sistematis, mulai dari pemilihan data hingga evaluasi model. Hasil penelitian menunjukkan bahwa RFR memberikan akurasi prediksi yang lebih tinggi dibandingkan regresi linear dalam memprediksi performa produk teh. Temuan ini bermanfaat bagi pelaku industri teh, termasuk produsen dan pemasar, dalam merumuskan strategi pemasaran yang lebih efektif berdasarkan data yang akurat. Penelitian ini juga merekomendasikan penelitian lebih lanjut untuk mengeksplorasi variabel lain yang mungkin mempengaruhi performa penjualan produk teh.

Kata Kunci: Performa Produk, Prediksi, *Regresi Linear*, *Random Forest*, Teh

I. PENDAHULUAN

Dunia bisnis mengalami perubahan signifikan di tengah pesatnya perkembangan teknologi dan transformasi digital yang mengharuskan perusahaan terus berinovasi demi mempertahankan dan meningkatkan persaingan (Hapriyanto, 2024). Dalam dunia bisnis yang semakin kompetitif, terutama dalam sektor pertanian dan perkebunan, kemampuan untuk memprediksi performa produk menjadi sangat krusial. Terlebih dalam konteks produk teh yang terkenal di Indonesia, berbagai faktor seperti harga jual, strategi pemasaran, dan volume produksi dapat mempengaruhi hasil penjualan secara signifikan. Untuk memahami pengaruh faktor-faktor tersebut dan untuk meningkatkan akurasi proyeksi penjualan, penggunaan metode analisis yang tepat sangat diperlukan. Oleh karena itu, penerapan algoritma prediktif seperti regresi linier dan Random Forest Regressor (RFR) menjadi sangat relevan dalam konteks ini.

Penelitian ini bertujuan untuk memprediksi performa teh dengan menilai penilaian konsumen melalui rating dan ulasan bintang. Dalam mencapai tujuan ini, penelitian menggunakan

model prediksi *Linear Regression* dan *Random Forest Regressor*. *Linear Regression* berfungsi mengidentifikasi pengaruh linier variabel seperti harga, jumlah ulasan, dan ukuran produk terhadap skor atau kualitas rata-rata produk. Sementara itu, *Random Forest Regressor* digunakan sebagai metode *ensemble learning* untuk menangkap hubungan yang lebih kompleks antara variabel bebas dan target, sehingga diharapkan dapat meningkatkan akurasi prediksi dengan mengenali pola preferensi konsumen yang beragam

Regresi linier, sebagaimana diungkapkan oleh (Andia, Kaslani, Permata, & Handayani, 2024) memiliki kemampuan untuk menjelaskan hubungan linear antara variabel yang merupakan bagian penting dalam menganalisis performa penjualan produk teh. Dengan pendekatan ini, peneliti dapat mengidentifikasi hubungan yang kuat antara variabel yang berbeda, sehingga memungkinkan pemangku kepentingan untuk membuat keputusan yang lebih tepat berdasarkan data historis. Ini terutama berguna dalam memahami bagaimana harga dan promosi mempengaruhi volume penjualan, yang merupakan kunci untuk strategi pemasaran yang sukses.

Di sisi lain, *Random Forest Regressor* menunjukkan potensi yang luar biasa dalam menangani dataset yang memiliki isi yang sangat banyak. Penelitian yang dilakukan oleh (Taihuttu, Sitanggang, & Syaufina, 2024) menegaskan keunggulan algoritma ini dalam memodelkan data dengan variabilitas tinggi tanpa terjebak dalam masalah *overfitting*. Dengan pengaturan *hyperparameter* yang baik, RFR dapat menghasilkan prediksi yang akurat, bahkan ketika data yang digunakan sangat bervariasi dan tidak teratur. Dalam konteks industri teh, di mana banyak variabel eksternal yang dapat mempengaruhi hasil, penggunaan RFR menjadi sangat penting untuk memastikan proyeksi yang lebih bagus.

Penelitian ini akan membandingkan regresi linier dan *random forest regressor* dalam konteks penjualan teh, regresi linier mungkin lebih tepat digunakan untuk analisis pengaruh variabel dan interpretasi hubungan, sedangkan *Random Forest Regressor* lebih unggul untuk tujuan peramalan penjualan dan pengambilan keputusan berbasis prediksi.

II. STUDI LITERATUR

Penelitian oleh (Hurifiani, Purnamasari, & Ali, 2024) menggunakan regresi linear berganda untuk meramalkan hasil panen padi di Kabupaten Cirebon, menghasilkan evaluasi dengan R^2 sebesar 0,99, yang menunjukkan performa prediksi yang sangat baik [1]. Selain itu, Hurifiani et al. menerapkan regresi linear sederhana untuk memprediksi penjualan Alat Tulis Kantor (ATK), dengan hasil evaluasi RMSE sebesar 7802.802 dan kesalahan relatif 21,08%, menunjukkan model yang cukup akurat meskipun menggunakan metode yang sederhana.

Di sisi lain, *Random Forest Regressor* (RFR) terbukti efektif dalam menangani data dengan kompleksitas tinggi dan variasi yang tidak teratur. Penelitian oleh (Taihuttu, Sitanggang, & Syaufina, 2024) menggunakan RFR untuk memprediksi kelembaban tanah di lahan gambut, mencapai akurasi tinggi dengan R^2 sebesar 0,914. Penelitian ini membuktikan keunggulan RFR dalam mengurangi *overfitting* dan menangani outliers. Selanjutnya, (Irfan Rasyid Muchtar; Afyati, 2024) melakukan perbandingan antara regresi linear dan *Random Forest* untuk prediksi harga beras premium, di mana *Random Forest* menghasilkan akurasi 98,69%, lebih tinggi dibandingkan dengan *Linear Regression* yang mencapai 95,08%. Ini menunjukkan efektivitas RFR dalam menangani data non-linear dan variabel kompleks.

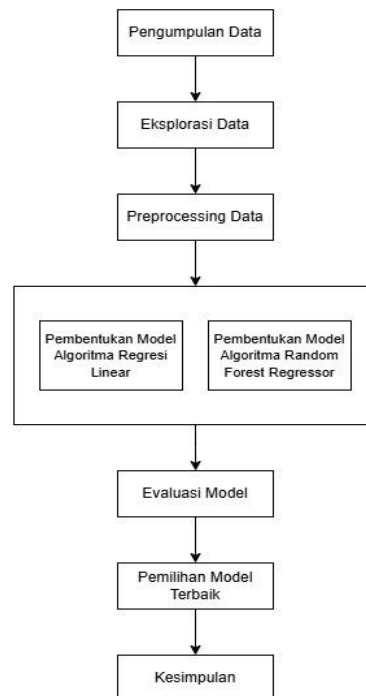
(Taihuttu, Sitanggang, & Syaufina, 2024) menemukan bahwa kelembaban tanah di lahan gambut memiliki keterkaitan yang kuat dengan kejadian kebakaran hutan dan lahan, di mana kelembaban tanah yang rendah meningkatkan risiko kekeringan yang berpotensi memicu kebakaran. Algoritma *Random Forest Regressor* (RFR) dipilih dalam penelitian ini karena kemampuannya untuk mendeteksi pola data yang kompleks dan non-linear. RFR juga unggul dalam memprediksi nilai dengan akurasi tinggi serta mampu mengatasi masalah *overfitting* dan outlier. Hasil penelitian menunjukkan bahwa model RFR yang telah dioptimalkan parameternya

mampu mencapai akurasi prediksi yang tinggi, dengan nilai R^2 sebesar 0,914 dan tingkat kesalahan yang rendah (MAE sebesar 0,0081 dan RMSE sebesar 0,0271).

(Hidayat, et al., 2025)Setiawan et al. meneliti penggunaan regresi linear dan RFR dalam memprediksi penjualan produk di supermarket. Meskipun RFR efektif untuk data kompleks, regresi linear juga mampu memberikan performa optimal dengan MAE sebesar 2.4899 dan R^2 sebesar 0.99.

III. METODE PENELITIAN

Alur penelitian dalam memprediksi performa produk teh menggunakan algoritma Random Forest Regressor dan Regresi Linier terdiri dari beberapa tahapan pada gambar 1.



Gambar 1. Alur Penelitian

Tahap awal ini berfokus pada pengumpulan data yang akan menjadi dasar bagi seluruh analisis. (Ginantra, et al., 2021)Seleksi atau pemilihan data perlu dilakukan dari himpunan data operasional sebelum tahapan penggalian informasi dalam *Knowledge Discovery in Databases* dimulai. Dataset yang digunakan diperoleh dari platform e-commerce iHerb, berisi data mengenai produk teh. Data ini mencakup beragam faktor yang berperan dalam performa produk teh, baik dari segi harga, ulasan, maupun kualitas. Dengan kumpulan data yang lengkap ini, penelitian dapat menelaah pengaruh berbagai variabel terhadap performa produk teh di pasaran.

Eksplorasi data bertujuan untuk memahami karakteristik dari setiap variabel sebelum memulai pemodelan. Tahap ini mencakup analisis statistik deskriptif, seperti menghitung nilai rata-rata, median, minimum, maksimum, dan standar deviasi dari variabel-variabel seperti `product_price`, `product_rate`, dan `reviews_number`. Visualisasi data dalam bentuk histogram dan box plot juga digunakan untuk melihat pola dan distribusi variabel, seperti harga produk dan rating.

Selain itu, korelasi antar variabel dianalisis untuk mengetahui apakah terdapat hubungan signifikan antar faktor. Sebagai contoh, korelasi antara `product_price` dan `product_rate` dapat menunjukkan apakah harga memengaruhi rating produk. Hubungan antara `reviews_number` dan `product_rate` juga akan diperiksa untuk melihat apakah rating yang lebih tinggi menghasilkan

lebih banyak ulasan. Eksplorasi ini memberikan pemahaman awal tentang pola yang ada dalam data dan menjadi dasar untuk preprocessing dan pemilihan fitur yang digunakan dalam model.

Setelah eksplorasi, data akan melalui tahap preprocessing dengan tujuan dari preprocessing ini adalah untuk meningkatkan kualitas output model dengan melakukan cleaning pada data, seperti memperbaiki atau menghapus data yang rusak atau tidak relevan untuk mempersiapkannya agar sesuai dengan model prediksi. Langkah-langkah preprocessing yang dilakukan meliputi pemilihan data dan atribut, pembersihan data, serta pemeriksaan seluruh nilai pada setiap atribut yang digunakan untuk mencegah ketidakseimbangan pada data (Schröe, Kruse, & Gómez, 2021).

Model Regresi Linear akan dibangun untuk memprediksi performa produk teh. (Ramelan & Hakim, 2025) Regresi Linear adalah model yang sederhana dan mudah diinterpretasikan yang berguna untuk menemukan hubungan linear antara variabel prediktor dengan variabel target, yaitu `product_rate` atau `reviews_number` (Sarah, Kurniadi, & Warsini, 2023). Model ini akan menggunakan berbagai variabel seperti `product_price`, `product_size`, distribusi ulasan (`one_star` hingga `five_star`), dan `product_quality`. Model Regresi Linear akan dilatih untuk memahami seberapa besar pengaruh setiap variabel terhadap performa produk teh. Hasil dari model ini memberikan wawasan tentang kontribusi variabel-variabel yang ada terhadap prediksi performa.

Model regresi linear sederhana dinyatakan dalam bentuk:

$$Y = \beta_0 + \beta_1 X + \epsilon \text{ (Asrirawan, Permata, \& Fausan, 2022)}$$

Pada model ini, diasumsikan bahwa hubungan antara X dan Y bersifat linear dalam parameter. Pemahaman tentang model ini adalah dasar yang penting untuk menyusun hipotesis mengenai pengaruh variabel X terhadap Y.

Model kedua yaitu Random Forest Regressor merupakan algoritma ensemble berbasis pohon keputusan. (Fauzi, Maulidah, Supriyadi, Nalatissifa, & Diantika, 2025) Algoritma ini menggabungkan hasil dari beberapa pohon keputusan untuk menghasilkan prediksi akhir, dan mampu menangkap hubungan non-linear antar variabel. Random forest adalah metode machine learning yang menggunakan konsep supervised dalam membangun kelas classifier (Saadah & Salsabila, 2021). Random Forest Regressor dilatih dengan variabel yang sama seperti Regresi Linear, yaitu `product_price`, `product_size`, distribusi ulasan, dan `product_quality`. Model ini diharapkan dapat mengidentifikasi pola yang lebih kompleks dalam data. Karena tahan terhadap overfitting, model ini sering kali lebih akurat dan stabil dibandingkan model linear, terutama jika data menunjukkan pola non-linear. Tuning parameter pada Random Forest Regressor dilakukan untuk memperoleh konfigurasi model yang optimal dalam memprediksi. Parameter yang disesuaikan meliputi jumlah pohon (`n_estimators`), kedalaman maksimum pohon (`max_depth`), jumlah minimum sampel untuk pemisahan node (`min_samples_split`), jumlah minimum sampel pada daun (`min_samples_leaf`), serta jumlah fitur yang dipertimbangkan pada setiap pemisahan (`max_features`). Proses tuning dilakukan menggunakan teknik validasi silang guna memastikan model memiliki kemampuan generalisasi yang baik dan tidak mengalami overfitting.

Tahap Evaluasi Model bertujuan untuk menilai performa kedua model dalam memprediksi performa produk teh. Metrik evaluasi yang digunakan mencakup, Mean Absolute Error (MAE) yaitu Rata-rata selisih absolut antara nilai prediksi dan nilai aktual, Mean Squared Error (MSE) adalah Rata-rata dari kuadrat kesalahan prediksi, memberi bobot lebih besar pada kesalahan yang lebih besar., serta Koefisien Determinasi (R^2) untuk Mengukur seberapa baik model menjelaskan variabilitas dari variabel target dibandingkan dengan rata-rata nilai target (Putra & Suhartana, 2025). R^2 berkisar dari 0 hingga 1, dan semakin mendekati 1 berarti model semakin baik dalam menjelaskan variasi data (Fitri, 2023).

Evaluasi dilakukan pada data pengujian untuk memastikan bahwa model tidak mengalami overfitting. Berdasarkan hasil evaluasi, model dengan performa terbaik akan dipilih sebagai model yang direkomendasikan. Selain akurasi, kemudahan interpretasi juga akan dipertimbangkan. Jika Random Forest lebih akurat tetapi terlalu kompleks, Regresi Linear mungkin tetap menjadi pilihan karena lebih mudah dimengerti dan diinterpretasikan. Pemilihan model terbaik akan mempertimbangkan kebutuhan dan tujuan analisis dari penelitian ini.

Ini termasuk penjelasan tentang faktor-faktor utama yang memengaruhi performa produk teh, seperti apakah harga atau rating lebih berpengaruh dalam menarik perhatian konsumen. Kesimpulan juga akan memberikan perbandingan antara Regresi Linear dan Random Forest dalam memprediksi performa produk teh, serta memberikan rekomendasi bagi perusahaan atau pelaku industri terkait strategi pemasaran dan pengembangan produk berdasarkan hasil prediksi yang dihasilkan model.

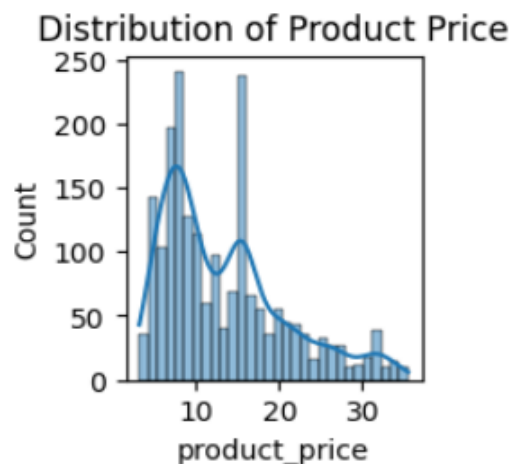
IV. HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini memuat informasi mengenai performa produk teh yang dijual di platform iHerb, dengan total 2.148 data produk. Setiap produk memiliki variabel-variabel seperti product_id, product_name, product_company, product_flavour, product_size, product_price, product_rate, reviews_number, serta kualitas produk dalam kategori product_quality. Data ini disajikan lengkap dan tidak ada nilai yang hilang, sehingga analisis lebih lanjut dapat dilakukan dengan data bersih, seperti yang dirangkum pada Tabel 1. berikut:

Tabel 1. Ringkasan tipe data dan ketiadaan missing values.

Kolom	Non-Null Count	Tipe Data
product_id	2148	int64
product_name	2148	object
product_company	2148	object
product_flavour	2148	object
product_size	2148	object
product_price	2148	float64
product_rate	2148	float64
reviews_number	2148	int64
product_quality	2148	object

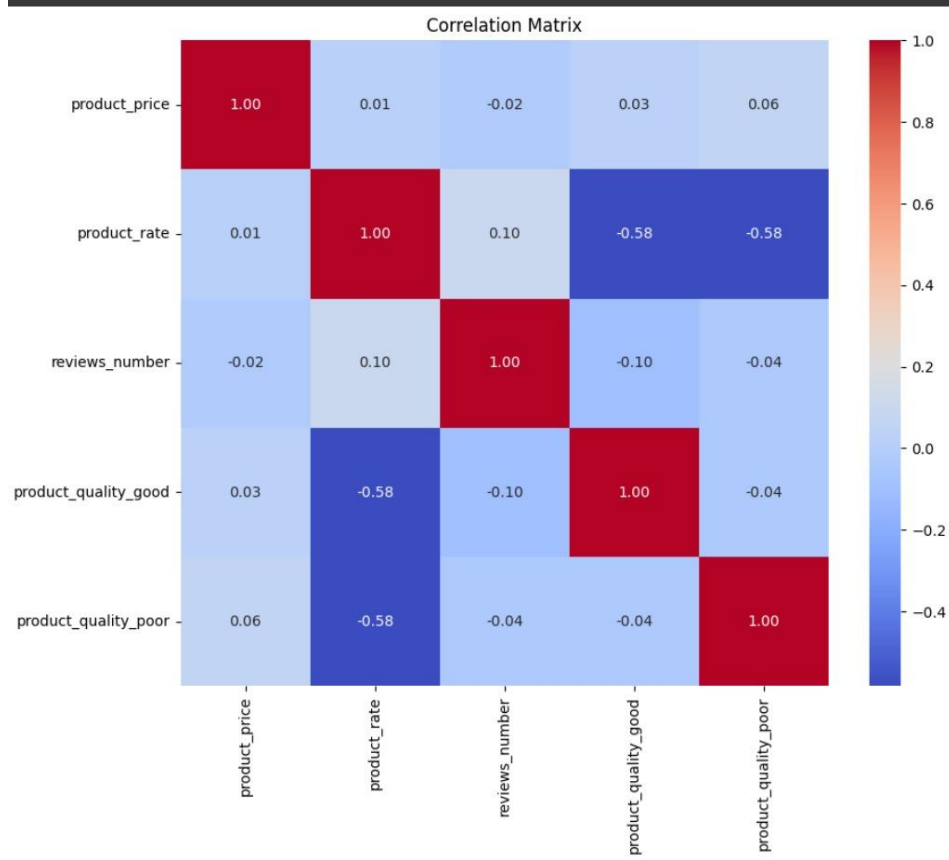
Rata-rata harga produk teh pada platform ini adalah \$20.35, dengan harga terendah \$5.50 dan tertinggi \$80.00. Distribusi harga ini ditunjukkan pada Gambar 4.1.2, yang memperlihatkan



bahwa sebagian besar produk teh berada di rentang harga antara \$10 hingga \$30, mencerminkan tren harga yang umum di pasar teh online.

Gambar 2. Distribusi harga product teh

Setelah preprocessing, dilakukan analisis korelasi untuk memahami hubungan antar fitur, terutama antara product_price, product_rate, dan reviews_number, serta variabel dummy dari product_quality. Korelasi ini divisualisasikan dengan heatmap untuk mempermudah interpretasi. Berikut adalah tabel korelasi antar variabel:



Gambar 3. Korelasi antar variable

Gambar di atas menampilkan matriks korelasi antara beberapa variabel terkait produk, yaitu product_price, product_rate, reviews_number, product_quality_good, dan product_quality_poor. Matriks korelasi menunjukkan seberapa kuat hubungan antara dua variabel, yang diukur dengan koefisien korelasi. Nilai korelasi berkisar dari -1 hingga 1, di mana nilai 1 menunjukkan korelasi positif sempurna (kedua variabel meningkat bersama), -1 menunjukkan korelasi negatif sempurna (satu variabel meningkat, sementara yang lain menurun), dan 0 menunjukkan tidak ada hubungan antara kedua variabel. Dari gambar di atas, tampak korelasi negatif yang cukup kuat antara product_rate dengan product_quality_good (-0.58), yang menunjukkan bahwa semakin tinggi kualitas yang dianggap baik, nilai produk (product_rate) cenderung menurun. Selain itu, product_rate juga memiliki korelasi negatif yang serupa dengan product_quality_poor, menunjukkan pola yang mungkin terkait dengan persepsi kualitas produk. Korelasi positif atau negatif yang kuat ditandai dengan warna merah (positif) atau biru tua (negatif), sedangkan warna biru muda menandakan korelasi yang lemah atau mendekati nol.

Perbandingan untuk kedua model dalam bentuk DataFrame yang menyajikan hasil evaluasi dari Regresi Linear dan Random Forest Regressor dapat dilihat pada tabel berikut:

Tabel 2. Perbandingan model

Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	R ² Score
Linear Regression	3.828×10^{-14}	3.522×10^{-27}	1.0
Random Forest	0.00064	6.034×10^{-5}	0.9997

Dari tabel ini, terlihat bahwa Linear Regression memiliki performa evaluasi yang sedikit lebih unggul dibandingkan Random Forest dalam hal MAE, MSE, dan R² Score. Nilai-nilai ini menegaskan bahwa model Linear Regression memberikan hasil yang lebih baik dan lebih akurat pada dataset ini, menjadikannya pilihan model terbaik di antara keduanya



Gambar 4. Scatter plot of product price vs product rate

Gambar di atas menunjukkan scatter plot yang menggambarkan hubungan antara product_price (harga produk) pada sumbu X dan product_rate (nilai produk) pada sumbu Y. Scatter plot juga menggunakan skema warna untuk menunjukkan tingkat nilai produk, di mana warna ungu menandakan nilai yang lebih rendah dan warna kuning menandakan nilai yang lebih tinggi. Dari gambar ini, terlihat bahwa sebagian besar produk dengan rentang harga yang bervariasi memiliki nilai yang tinggi, yakni di kisaran 4 hingga 5, yang ditunjukkan oleh titik-titik berwarna kuning dan hijau. Namun, terdapat beberapa produk dengan harga tinggi yang memiliki nilai rendah, begitu pula sebaliknya beberapa produk dengan harga rendah memiliki nilai tinggi. Secara keseluruhan, scatter plot menunjukkan bahwa tidak ada hubungan linier yang kuat antara harga dan nilai produk, yang ditunjukkan oleh persebaran data yang cukup acak pada harga-harga produk yang berbeda. Produk-produk dengan harga menengah tampaknya mendapatkan nilai yang lebih tinggi, sementara harga yang lebih ekstrem, baik tinggi maupun rendah, cenderung memiliki nilai yang lebih rendah.

V. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa dataset yang digunakan, yang memuat informasi mengenai performa produk teh di platform iHerb dengan 2.148 data produk, tidak memiliki nilai yang hilang, sehingga memungkinkan untuk dilakukan analisis tanpa kendala terkait data yang hilang. Distribusi harga produk teh rata-rata sebesar \$20.35, dengan harga terendah \$5.50 dan tertinggi \$80.00, menunjukkan bahwa sebagian besar produk teh berada di rentang harga \$10 hingga \$30. Korelasi antarvariabel menunjukkan hubungan yang menarik, yaitu korelasi positif sebesar 0.35 antara harga (`product_price`) dan rating produk (`product_rate`), yang mengindikasikan bahwa produk dengan harga lebih tinggi cenderung memiliki rating yang sedikit lebih baik. Sebaliknya, terdapat korelasi negatif -0.58 antara `product_rate` dan `product_quality_good`, yang menunjukkan bahwa produk yang dianggap berkualitas baik justru memiliki rating sedikit lebih rendah. Dalam tahap pemodelan, dua model diterapkan, yaitu Regresi Linear dan Random Forest Regressor. Evaluasi model menunjukkan bahwa Regresi Linear memiliki performa lebih unggul dengan nilai R^2 sebesar 1.0, Mean Absolute Error (MAE) sebesar 3.828×10^{-14} , dan Mean Squared Error (MSE) sebesar 3.522×10^{-27} , menjadikannya model terbaik untuk dataset ini. Model Random Forest juga menunjukkan performa yang sangat baik dengan nilai R^2 sebesar 0.9997, MAE sebesar 0.00064, dan MSE sebesar 6.034×10^{-5} , namun performanya sedikit di bawah Regresi Linear. Selain itu, scatter plot antara `product_price` dan `product_rate` menunjukkan bahwa tidak ada hubungan linier yang kuat antara harga dan rating produk, meskipun sebagian besar produk dengan harga menengah hingga tinggi memiliki rating yang baik di kisaran 4 hingga 5. Kesimpulannya, model Regresi Linear lebih unggul dalam memprediksi rating produk teh di platform ini, dan hasil ini mengindikasikan bahwa harga tidak sepenuhnya menentukan rating, melainkan ada faktor lain yang mempengaruhi persepsi konsumen terhadap kualitas dan rating produk. Penelitian ini memberikan wawasan bagi platform atau penjual produk teh untuk menyusun strategi terkait harga dan kualitas produk sesuai dengan ekspektasi konsumen..

VI. REFERENSI

- ANDIA, R., KASLANI, PERMATA, S. E., & HANDAYANI, T. (2024). PERAMALAN HASIL PANEN PADI KABUPATEN CIREBON MENGGUNAKAN ALGORITMA REGRESI LINEAR BERGANDA. *JATI (JURNAL MAHASISWA TEKNIK INFORMATIKA)*.
- Asrirawan, Permata, S. U., & Fausan, M. I. (2022). Pendekatan Univariate Time Series Modelling untuk Prediksi Kuartalan Pertumbuhan Ekonomi Indonesia Pasca Vaksinasi COVID-19. *JAMBURA JOURNAL OF MATHEMATICS*, 86-103.
- Fauzi, A., Maulidah, N., Supriyadi, R., Nalattissifa, H., & Diantika, S. (2025). Prediksi Harga Properti Di Indonesia Menggunakan Algoritma Random Forest. *Journal of Artificial Intelligence and Digital Business(RIGGS)*.
- Fitri, E. (2023). Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah. *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, 58-64.
- Ginantra, N. L., Arifah, F. N., Wijaya, A. H., Septarini, R. S., Ahmad, N., Ardiana, D. P., . . . Neg, E. S. (2021). *Data Mining dan Penerapan Algoritma*. Medan: Kita Menulis.
- Hapriyanto, A. R. (2024). Multidisciplinary Science Strategi Inovatif dalam Meningkatkan Daya Saing Bisnis di Era Digital. *Nusantara Journal of Multidisciplinary Science*.
- Hidayat, R., Saputra, H. T., Husnah, M., Nabila, Hidayatullah, M. B., Nazhmi, M. N., . . . Rana, A. (2025). Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produk di Supermarket. *Jurnal Sistem Informasi dan Sistem Komputer*.

-
- Hurifiani, A., Purnamasari, A. I., & Ali, I. (2024). PENERAPAN ALGORITMA REGRESI LINEAR UNTUK PREDIKSI PENJUALAN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 266-273.
- Irfan Rasyid Muchtar; Afiyati. (2024). Comparison of Linear Regression and Random Forest Algorithms for Premium Rice Price Prediction (Case Study: West Java). *JIST*.
- Putra, I. M., & Suhartana, I. K. (2025). Implementasi Algoritma Random Forest Regression dalam Sistem Prediksi Harga Rumah di Jabodetabek . *Jurnal Nasional Teknologi Informasi dan Aplikasinya*.
- Ramelan, F. A., & Hakim, L. (2025). Perbandingan Prediksi terhadap Peningkatan Jumlah Pelanggan Iconnet dengan Algoritma Regresi Linear dan Random Forest pada Wilayah Jabodetabek dan Banten. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 790-8-1.
- Saadah, S., & Salsabila, H. (2021). Prediksi Harga Bitcoin Menggunakan Metode Random Forest. *Jurnal Politeknik Caltex Riau*, 24-32.
- Sarah, A. M., Kurniadi, B., & Warsini, E. (2023). IMPLEMENTASI METODE REGRESI LINEAR DALAM MEMPREDIKSI PENYAKIT ANEMIA SECARA DINI. *JURNAL TEKNISI (Jurnal Teknologi Komputer dan Sistem Informasi)*, 12-23.
- Schröe, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 526-534.
- Taihuttu, H. Y., Sitanggang, I. S., & Syaufina, L. (2024). SOIL MOISTURE PREDICTION MODEL IN PEATLAND USING RANDOM FOREST REGRESSOR. *BAREKENG: Journal of Mathematics and Its Applications*, 2505–2516.