

PEMANFAATAN DATA MINING UNTUK PREDIKSI HARGA SMARTPHONE MENGGUNAKAN RANDOM FOREST REGRESSOR

Daffa Alif Ruriyanto ¹, Fauzan Abhip Raya ², and Reffiano Siswoyo³

Program Studi Sistem Informasi, Universitas Pamulang, Tangerang Selatan, Indonesia, 15310
e-mail: ¹ notdaffa@gmail.com, ² fauzanabhipraya14@gmail.com, ³ reffzhii@gmail.com

Abstract

The rapid development of smartphone technology has led to an increasing variety of devices available in the market with diverse specifications and prices. This condition often makes prospective buyers experience difficulties in determining a device that matches their needs and budget. Therefore, a data analysis approach is needed to predict smartphone prices based on their technical specifications. This study aims to build a smartphone price prediction model using the Random Forest Regressor method by utilizing a dataset containing various device specification information such as rating, specification score, RAM capacity, battery capacity, screen size, camera, and internal memory. This research applies the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology which includes the stages of business understanding, data understanding, data preparation, modeling, and evaluation. The analysis process is carried out using the Python programming language with supporting libraries such as Pandas, NumPy, and Scikit-learn. The results show that the Random Forest Regressor model is able to identify the relationship between device specifications and smartphone prices, so it can be used to estimate smartphone prices more objectively.

Keywords: Data Mining, Random Forest Regressor, CRISP-DM, Smartphone, Price Prediction

Abstrak

Perkembangan teknologi smartphone yang pesat menyebabkan semakin banyaknya variasi perangkat dengan spesifikasi dan harga yang beragam di pasar. Kondisi ini seringkali membuat calon pembeli mengalami kesulitan dalam menentukan perangkat yang sesuai dengan kebutuhan dan anggaran yang dimiliki. Oleh karena itu, diperlukan suatu pendekatan analisis data untuk memprediksi harga smartphone berdasarkan spesifikasi teknisnya. Penelitian ini bertujuan membangun model prediksi harga smartphone menggunakan metode Random Forest Regressor dengan memanfaatkan dataset yang berisi berbagai informasi spesifikasi perangkat seperti rating, skor spesifikasi, RAM, kapasitas baterai, ukuran layar, kamera, dan memori internal. Penelitian ini menggunakan metodologi CRISP-DM yang meliputi tahapan pemahaman masalah, pemahaman data, persiapan data, pemodelan, dan evaluasi. Proses analisis dilakukan menggunakan Python dengan pustaka Pandas, NumPy, dan Scikit-learn. Hasil penelitian menunjukkan bahwa model Random Forest Regressor mampu mengidentifikasi hubungan antara spesifikasi perangkat dan harga smartphone sehingga dapat digunakan untuk memperkirakan harga perangkat secara lebih objektif.

Kata Kunci: Data Mining, Random Forest Regressor, CRISP-DM, Smartphone, Price Prediction

1. PENDAHULUAN

Perkembangan teknologi smartphone yang pesat menyebabkan semakin banyaknya variasi perangkat dengan spesifikasi dan harga yang beragam di pasaran. Kondisi ini seringkali menimbulkan kesulitan bagi calon pembeli dalam menentukan perangkat yang sesuai dengan kebutuhan dan anggaran yang dimiliki. Banyaknya pilihan smartphone yang tersedia membuat pengguna tidak selalu dapat menilai apakah harga suatu perangkat telah sebanding dengan spesifikasi yang ditawarkan.

Permasalahan tersebut menunjukkan bahwa diperlukan suatu pendekatan yang mampu membantu pengguna dalam memahami hubungan antara spesifikasi smartphone dan harga perangkat secara lebih objektif. Salah satu pendekatan yang dapat digunakan adalah data mining, yaitu proses untuk mengekstraksi pola atau informasi penting dari sekumpulan data dalam jumlah besar. Data mining dapat digunakan untuk mengekstraksi pola dari data [1]. Dengan memanfaatkan data yang tersedia, proses pengambilan keputusan dapat dilakukan secara lebih sistematis dan berbasis data. Dalam penerapannya, data mining sering dikombinasikan dengan teknik machine learning, yang memungkinkan sistem untuk mempelajari pola dari data historis dan menggunakannya untuk melakukan prediksi terhadap data baru [2]. Pendekatan ini telah banyak digunakan dalam berbagai bidang untuk membangun model prediksi yang akurat dan efisien.

Salah satu metode yang umum digunakan dalam machine learning untuk permasalahan prediksi adalah Random Forest Regressor, yang termasuk dalam kategori supervised learning. Metode ini digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen, sehingga dapat digunakan untuk memperkirakan nilai suatu variabel berdasarkan variabel lainnya [3].

Dataset yang digunakan dalam penelitian ini berisi berbagai informasi terkait spesifikasi smartphone, seperti rating, skor spesifikasi (spec score), kapasitas RAM, kapasitas baterai, ukuran layar, kamera, memori internal, dukungan jaringan, prosesor, serta harga perangkat. Dengan memanfaatkan data tersebut, model machine

learning dapat dibangun untuk memprediksi harga smartphone berdasarkan spesifikasi teknis yang dimilikinya.

Selain itu, terdapat pendekatan lain dalam data mining yang juga digunakan untuk menganalisis pola data. Selain itu, terdapat pendekatan lain dalam data mining yang juga digunakan untuk menganalisis pola data, yaitu algoritma Apriori. Algoritma ini digunakan untuk menemukan hubungan atau pola asosiasi antar item dalam suatu dataset melalui pembentukan kombinasi itemset dan aturan asosiasi berdasarkan nilai support dan confidence. Penelitian yang dilakukan oleh Karimah dan Pahira (2022) menunjukkan bahwa metode Apriori dapat digunakan untuk menganalisis pola penjualan dengan mengidentifikasi kombinasi item yang sering muncul secara bersamaan, sehingga dapat dimanfaatkan dalam mendukung pengambilan keputusan berbasis data [4].

Namun, metode Apriori lebih berfokus pada pencarian pola hubungan antar data dan tidak digunakan untuk memprediksi nilai numerik secara langsung. Oleh karena itu, dalam penelitian ini digunakan metode Random Forest Regressor yang bertujuan untuk memodelkan hubungan antara spesifikasi smartphone dan harga perangkat sehingga dapat menghasilkan prediksi harga secara kuantitatif.

Penelitian ini bertujuan untuk membangun model prediksi harga smartphone serta menganalisis hubungan antara spesifikasi perangkat dengan harga yang terbentuk di pasaran. Hasil dari penelitian ini diharapkan dapat memberikan gambaran mengenai faktor-faktor yang mempengaruhi harga smartphone, serta membantu pengguna dalam menentukan pilihan perangkat secara lebih rasional, objektif, dan berbasis data.

2. METODE

Penelitian ini menggunakan pendekatan data mining untuk membangun model prediksi harga smartphone berdasarkan spesifikasi teknis perangkat. Pendekatan ini dipilih karena mampu mengekstraksi pola dan hubungan yang tersembunyi dalam dataset sehingga dapat

Selain melakukan eksplorasi nilai distribusi, pada tahap data understanding ini juga dilakukan identifikasi data pencilan (outlier) menggunakan metode boxplot pada fitur-fitur numerik utama. Deteksi ini bertujuan untuk mengevaluasi variabilitas data serta potensi adanya nilai ekstrem

Regressor saat proses pemodelan. Hasil visualisasi deteksi pencilan dapat dilihat pada Gambar Berikut:

Gambar 2. Visualisasi (Outlier) pada Fitur Numerik: (a) Spec_score, (b) Rating.

Berdasarkan Gambar 2 (a), pada fitur Spec_score terdeteksi adanya sejumlah data pencilan (outlier) yang berada jauh di bawah ambang batas nilai minimum. Sementara itu, fitur Rating pada Gambar 2 (b) menunjukkan persebaran data yang relatif bersih dari pencilan ekstrem di luar batas whisker. Temuan adanya pencilan pada variabel Spec_score ini menjadi acuan penting untuk penanganan dan pembersihan data pada tahap Data Preparation selanjutnya guna meningkatkan kualitas data dan performa model Random Forest Regressor.

c) Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini merupakan data spesifikasi smartphone yang mencerminkan karakteristik teknis perangkat yang beredar di pasar. Setiap baris data merepresentasikan satu perangkat smartphone, sedangkan setiap kolom menunjukkan atribut atau fitur yang menggambarkan spesifikasi perangkat tersebut. Dataset ini digunakan untuk menganalisis hubungan antara spesifikasi teknis smartphone dengan harga perangkat sebagai variabel target.

Dalam konteks data mining, dataset dengan struktur atribut seperti ini memungkinkan dilakukan analisis hubungan antar variabel untuk menemukan pola yang dapat digunakan dalam proses prediksi [9]. Dengan demikian, data yang digunakan tidak hanya berfungsi sebagai sumber informasi, tetapi juga sebagai dasar dalam membangun model machine learning yang mampu merepresentasikan hubungan antara spesifikasi perangkat dan harga smartphone secara sistematis.

d) Data Preparation

Tahap data preparation merupakan proses pengolahan data sebelum digunakan dalam pemodelan. Pada tahap ini dilakukan pembersihan data dengan menangani nilai yang hilang (missing values) menggunakan teknik imputasi. Selanjutnya, variabel kategorikal diubah menjadi bentuk numerik menggunakan teknik encoding

agar dapat diproses oleh algoritma machine learning. Selanjutnya dilakukan proses encoding pada variabel kategorikal dan penanganan missing value agar data dapat digunakan dalam proses pemodelan [10].

Selain itu, dataset dibagi menjadi data pelatihan (training data) dan data pengujian (testing data) menggunakan teknik train-test split untuk memastikan model dapat dievaluasi secara objektif dan menghindari overfitting. Tahapan ini merupakan bagian penting dalam proses data mining karena kualitas data sangat mempengaruhi hasil model yang dihasilkan.

e) Modeling

Tahap modeling dilakukan dengan membangun model prediksi menggunakan algoritma Random Forest Regressor. Algoritma ini digunakan untuk memodelkan hubungan antara variabel independen berupa spesifikasi smartphone dengan variabel dependen berupa harga perangkat. Random Forest Regressor merupakan algoritma ensemble learning yang bekerja dengan membangun banyak pohon keputusan (decision tree) dan menggabungkan hasil prediksi dari seluruh pohon tersebut untuk menghasilkan prediksi yang lebih akurat dan stabil. Metode ini mampu menangani hubungan yang kompleks antara variabel independen dan variabel dependen serta mengurangi risiko overfitting.[3].

f) Evaluation dan Interpretation of Results

Tahap evaluasi dilakukan untuk mengukur kinerja model dalam memprediksi harga smartphone. Evaluasi dilakukan menggunakan metrik seperti Mean Squared Error (MSE) dan koefisien determinasi (R^2) untuk mengetahui tingkat kesalahan serta kemampuan model dalam menjelaskan variasi data [11]. Metrik tersebut umum digunakan dalam model regresi untuk menilai seberapa baik hasil prediksi mendekati nilai aktual.

3. PEMBAHASAN

Bagian hasil dan pembahasan pada penelitian ini menyajikan hasil analisis data serta evaluasi model yang telah dibangun untuk memprediksi harga smartphone berdasarkan spesifikasi teknis perangkat. Pembahasan dilakukan secara bertahap,

dimulai dari analisis deskriptif dataset, visualisasi hubungan antar variabel, proses pemodelan menggunakan algoritma Random Forest Regressor, hingga evaluasi kinerja model dan implementasi solusi yang dihasilkan. Melalui tahapan ini, diharapkan dapat diperoleh pemahaman yang komprehensif mengenai hubungan antara spesifikasi smartphone dan harga perangkat di pasaran.

a) Analisis Deskriptif Dataset

Analisis deskriptif dilakukan untuk memahami karakteristik dasar dataset yang digunakan dalam penelitian. Berdasarkan hasil perhitungan statistik deskriptif, diperoleh bahwa variabel seperti kapasitas RAM memiliki rata-rata sekitar 6-8 GB dengan nilai maksimum mencapai 16 GB, sedangkan kapasitas baterai memiliki rata-rata sekitar 4500-5000 mAh. Variabel memori internal menunjukkan variasi yang cukup besar, mulai dari 32 GB hingga 512 GB.

Variabel harga sebagai target menunjukkan rentang nilai yang cukup luas, mulai dari kategori harga rendah hingga perangkat kelas premium. Hal ini menunjukkan bahwa dataset mencakup berbagai segmen pasar smartphone, sehingga model yang dibangun mampu mempelajari pola hubungan yang beragam antara spesifikasi dan harga perangkat.

b) Visualisasi Korelasi Antar Variabel

Visualisasi heatmap korelasi digunakan untuk mengidentifikasi hubungan antar variabel numerik dalam dataset yang digunakan dalam penelitian. Melalui heatmap ini dapat diketahui seberapa kuat hubungan antar variabel dengan melihat nilai koefisien korelasi yang berkisar antara -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan

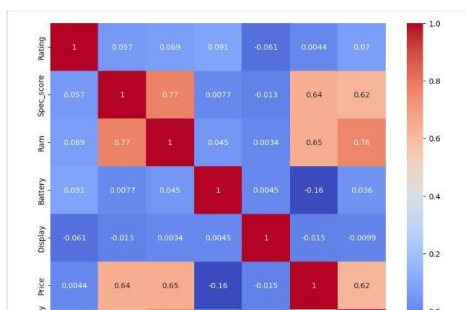
hubungan positif yang kuat, nilai mendekati 0 menunjukkan hubungan yang lemah, dan nilai mendekati -1 menunjukkan hubungan negatif.

Berdasarkan hasil visualisasi heatmap, terlihat bahwa variabel RAM dan Inbuilt_memory memiliki korelasi positif yang cukup kuat terhadap harga smartphone, dengan nilai korelasi masing-masing sekitar 0,64 dan 0,62. Hal ini menunjukkan bahwa semakin besar kapasitas RAM dan memori internal yang dimiliki oleh suatu perangkat, maka harga smartphone cenderung semakin tinggi. Kondisi ini dapat dipahami karena kapasitas memori yang lebih besar biasanya berkaitan dengan performa perangkat yang lebih baik serta kemampuan menjalankan aplikasi yang lebih kompleks.

Selain itu, variabel Spec_score juga menunjukkan hubungan positif terhadap harga dengan nilai korelasi sekitar 0,64. Variabel ini merepresentasikan skor keseluruhan spesifikasi perangkat, sehingga secara logis perangkat dengan nilai spesifikasi yang lebih tinggi akan memiliki harga yang lebih mahal di pasar. Nilai korelasi tersebut menunjukkan bahwa Spec_score memiliki hubungan yang cukup kuat terhadap harga smartphone dan dapat dijadikan salah satu indikator penting dalam proses prediksi harga. Namun, harga smartphone juga dipengaruhi oleh berbagai faktor lain yang tidak sepenuhnya tercermin dalam skor spesifikasi tersebut.

Variabel lain seperti kapasitas baterai dan ukuran display menunjukkan nilai korelasi yang relatif rendah terhadap harga smartphone. Hal ini mengindikasikan bahwa kedua faktor tersebut tidak menjadi penentu utama dalam variasi harga perangkat pada dataset yang digunakan. Meskipun demikian, baterai dan ukuran layar tetap memiliki peran dalam pengalaman penggunaan perangkat, namun pengaruhnya terhadap harga tidak sebesar komponen performa seperti RAM dan memori internal.

Sementara itu, variabel rating pengguna menunjukkan korelasi yang sangat rendah terhadap harga, bahkan mendekati nol. Hal ini menunjukkan bahwa tingkat kepuasan atau penilaian pengguna tidak secara langsung mempengaruhi harga smartphone dalam dataset yang dianalisis. Dengan kata lain, harga perangkat lebih banyak ditentukan

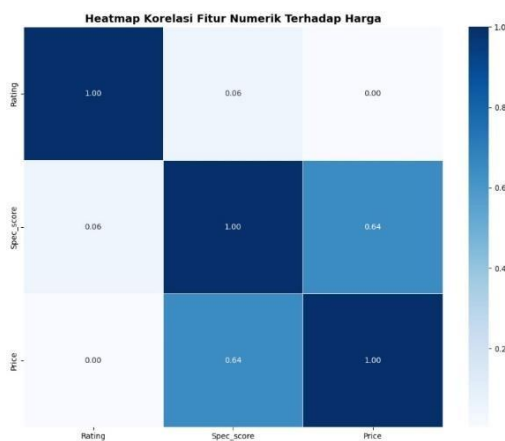


Gambar 3. Heatmap Korelasi Antar Variabel

oleh karakteristik teknis dan spesifikasi hardware dibandingkan dengan faktor subjektif yang berasal dari persepsi pengguna.

Secara keseluruhan, hasil analisis korelasi ini memberikan gambaran awal mengenai variabel-variabel yang memiliki hubungan paling signifikan terhadap harga smartphone. Informasi ini menjadi penting dalam tahap pemodelan karena membantu mengidentifikasi fitur yang berpotensi memberikan kontribusi besar dalam proses prediksi harga menggunakan metode machine learning.

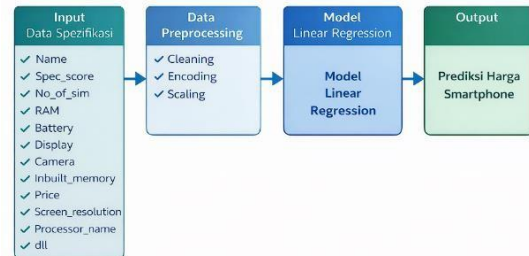
Untuk memahami hubungan antar variabel dalam dataset, dilakukan visualisasi menggunakan heatmap korelasi. Visualisasi ini digunakan untuk mengidentifikasi tingkat hubungan antara masing-masing variabel numerik terhadap harga smartphone.



Gambar 5. Heatmap Korelasi Fitur Numerik Terhadap Harga

Heatmap tersebut menampilkan korelasi antara fitur numerik dalam dataset terhadap harga, yaitu variabel Rating, Spec_score, dan Price. Korelasi ini menunjukkan seberapa kuat hubungan linear antarvariabel, dengan nilai berkisar antara -1 hingga 1. Nilai yang mendekati 1 menunjukkan hubungan positif yang kuat, nilai yang mendekati 0 menunjukkan hubungan yang sangat lemah atau hampir tidak ada hubungan, sedangkan nilai yang mendekati -1 menunjukkan hubungan negatif yang kuat.

Berdasarkan heatmap tersebut, terlihat bahwa Spec_score memiliki korelasi sebesar 0,64 terhadap Price, yang menunjukkan adanya



Gambar 4. Diagram Alur Sistem Prediksi Harga Smartphone

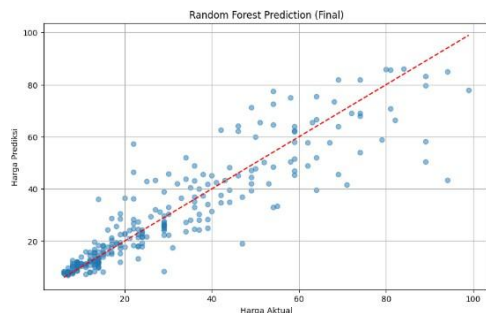
hubungan positif yang cukup kuat antara skor spesifikasi dengan harga. Hal ini mengindikasikan bahwa semakin tinggi nilai spesifikasi suatu produk, maka harga produk tersebut cenderung semakin tinggi. Dengan demikian, variabel Spec_score dapat dianggap sebagai salah satu faktor yang cukup berpengaruh dalam menentukan harga.

Sementara itu, variabel Rating memiliki korelasi yang sangat rendah terhadap Price, yaitu sebesar 0,00. Nilai ini menunjukkan bahwa rating hampir tidak memiliki hubungan langsung dengan harga, sehingga perubahan pada rating tidak secara signifikan memengaruhi perubahan harga dalam dataset tersebut. Selain itu, hubungan antara Rating dan Spec_score juga tergolong sangat lemah, dengan nilai korelasi sebesar 0,06, yang menunjukkan bahwa kedua variabel tersebut hampir tidak saling berkaitan secara linear.

Secara keseluruhan, heatmap ini menunjukkan bahwa Spec_score merupakan variabel yang paling berkaitan dengan harga dibandingkan dengan Rating. Oleh karena itu, dalam analisis atau pemodelan prediksi harga, variabel Spec_score kemungkinan memiliki kontribusi yang lebih signifikan, sedangkan variabel Rating mungkin memiliki pengaruh yang relatif kecil terhadap perubahan harga.

c) Hasil Permodelan Random Forest Regressor
Model prediksi dibangun menggunakan algoritma Random Forest Regressor untuk mempelajari hubungan antara variabel independen dan variabel dependen. Dataset dibagi menjadi data

pelatihan dan data pengujian dengan proporsi 80:20 menggunakan teknik train-test split. Model dilatih



Gambar 6. Perbandingan Harga Aktual vs Prediksi (Metode Random Forest Regressor) menggunakan data pelatihan dan diuji menggunakan data pengujian.

Hasil pemodelan menunjukkan bahwa model Random Forest Regressor mampu membentuk hubungan matematis antara spesifikasi smartphone dan harga perangkat. Hasil pemodelan menunjukkan bahwa variabel seperti RAM, memori internal, dan skor spesifikasi memberikan kontribusi yang signifikan dalam proses prediksi harga smartphone.

Untuk memberikan gambaran alur sistem prediksi harga smartphone yang dibangun, disajikan diagram alur sistem yang menunjukkan proses dari input data hingga menghasilkan output berupa prediksi harga.

Berdasarkan Gambar 5, sistem dimulai dengan memasukkan data spesifikasi smartphone sebagai input. Data tersebut kemudian melalui tahap preprocessing untuk memastikan kualitas data sebelum digunakan dalam proses pemodelan. Selanjutnya, data yang telah diproses dimasukkan ke dalam model Random Forest Regressor untuk menghasilkan prediksi harga smartphone. Output yang dihasilkan berupa estimasi harga perangkat yang dapat digunakan sebagai acuan dalam pengambilan keputusan.

d) Evaluasi Model

Evaluasi model dilakukan untuk mengukur kemampuan algoritma Random Forest Regressor dalam memprediksi harga smartphone berdasarkan spesifikasi perangkat yang digunakan sebagai variabel input. Pada tahap ini digunakan beberapa metrik evaluasi, yaitu Mean Squared Error (MSE),

Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2). Metrik-metrik tersebut digunakan untuk mengetahui tingkat kesalahan prediksi serta kemampuan model dalam menjelaskan variasi data harga smartphone. MSE dan RMSE digunakan untuk mengukur besarnya kesalahan prediksi, MAPE digunakan untuk mengetahui persentase rata-rata kesalahan prediksi terhadap nilai aktual, sedangkan R^2 digunakan untuk mengukur kemampuan model dalam menjelaskan hubungan antara variabel independen dan variabel dependen.

Berdasarkan hasil pengujian model, diperoleh nilai R^2 pada data training sebesar 0,9639 dan R^2 pada data testing sebesar 0,8165. Nilai R^2 pada data testing menunjukkan bahwa sekitar 81,65% variasi harga smartphone dapat dijelaskan oleh variabel spesifikasi yang digunakan dalam model, sedangkan sisanya dipengaruhi oleh faktor lain yang tidak termasuk dalam dataset. Nilai R^2 yang cukup tinggi ini menunjukkan bahwa model memiliki kemampuan prediksi yang baik.

Selain itu, nilai Root Mean Squared Error (RMSE) sebesar 9,50 menunjukkan bahwa rata-rata kesalahan prediksi model terhadap harga aktual berada pada kisaran sekitar 9,5 satuan harga pada dataset. Sementara itu, nilai Mean Absolute Percentage Error (MAPE) sebesar 20,79% menunjukkan bahwa rata-rata kesalahan prediksi model berada pada kisaran sekitar 20,79% dari nilai harga sebenarnya.

Visualisasi hubungan antara harga aktual dan harga prediksi juga ditunjukkan melalui grafik scatter plot. Pada grafik tersebut, garis diagonal merah menunjukkan kondisi prediksi ideal di mana nilai prediksi sama dengan nilai aktual. Sebagian besar titik data berada di sekitar garis tersebut, yang menunjukkan bahwa model mampu melakukan prediksi dengan cukup baik meskipun masih terdapat beberapa penyimpangan pada beberapa data.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model Random Forest yang digunakan memiliki performa yang cukup baik dalam memprediksi harga smartphone berdasarkan spesifikasi perangkat.

e) Implementasi Strategi Solusi

Model yang telah dibangun kemudian diimplementasikan sebagai solusi untuk membantu memprediksi harga smartphone berdasarkan spesifikasi perangkat. Dalam implementasinya, pengguna dapat memasukkan nilai spesifikasi seperti kapasitas RAM, baterai, ukuran layar, kamera, dan memori internal ke dalam model untuk memperoleh estimasi harga perangkat.

Hasil prediksi yang dihasilkan dapat digunakan sebagai acuan dalam menilai apakah harga suatu smartphone tergolong wajar dibandingkan dengan spesifikasi yang ditawarkan. Selain itu, model ini juga dapat digunakan untuk memperkirakan harga perangkat baru yang belum tersedia dalam dataset. Dengan demikian, model prediksi ini berpotensi dikembangkan menjadi sistem rekomendasi atau sistem pendukung keputusan dalam pemilihan smartphone.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode Random Forest Regressor dapat digunakan untuk membangun model prediksi harga smartphone berdasarkan spesifikasi teknis perangkat. Melalui pendekatan data mining dengan metodologi CRISP-DM, proses penelitian dapat dilakukan secara sistematis mulai dari tahap Business Understanding, Data Understanding, Data Preparation, Modeling, hingga Evaluation.

Hasil analisis menunjukkan bahwa beberapa atribut spesifikasi memiliki hubungan yang cukup kuat terhadap harga smartphone, terutama Spec_score, RAM, dan kapasitas memori internal. Semakin tinggi spesifikasi yang dimiliki suatu perangkat, maka harga smartphone cenderung semakin tinggi. Sebaliknya, variabel Rating menunjukkan hubungan yang relatif rendah terhadap harga sehingga kontribusinya dalam proses prediksi tidak terlalu signifikan.

Berdasarkan hasil evaluasi model, algoritma Random Forest Regressor mampu menghasilkan prediksi harga yang cukup baik dan mendekati nilai aktual. Hal ini menunjukkan bahwa hubungan

antara spesifikasi smartphone dan harga perangkat dapat direpresentasikan dengan baik menggunakan model Random Forest Regressor. Selain itu, visualisasi perbandingan harga aktual dan harga prediksi menunjukkan bahwa sebagian besar data berada di sekitar garis prediksi ideal, sehingga model mampu mengikuti pola harga smartphone yang terdapat pada dataset.

DAFTAR PUSTAKA

- [1] D. Marbun, F.; Zarlis, M.; Hartama, "Penerapan Data Mining untuk Prediksi Menggunakan Algoritma Regresi Linear," *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 567–573, 2021.
- [2] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper," *Indones. J. Comput. Inf. Technol.*, vol. 5, no. 1, 2020.
- [3] M. Sholeh, E. K. Nurnawati, and U. Lestari, "Penerapan Data Mining dengan Metode Random Forest Regressor untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 10–21, 2023, doi: 10.14421/jiska.2023.8.1.10-21.
- [4] M. Karimah and W. Pahira, "Penerapan Algoritma Apriori Pada Prediksi Penjualan Tanaman Hias Bromelia (Studi Kasus: Karimah Flora)," *Maret*, vol. 7, no. 1, pp. 2622–4615, 2021, [Online]. Available: <http://openjournal.unpam.ac.id/index.php/informatika68>
- [5] D. S. Panggabean, E. Buulolo, and N. Silalahi, "Penerapan Data Mining untuk Prediksi Menggunakan Metode Random Forest Regressor," *J. Media Inform. Budidarma*, vol. 4, no. 3, pp. 604–610, 2020.
- [6] R. A. Siregar, "Penerapan Metode CRISP-DM pada Data Mining untuk Analisis Data," *J. Ristek*, vol. 6, no. 2, 2022.
- [7] T. Hidayat and R. Sari, "Implementasi Data Mining Menggunakan Metode CRISP-DM untuk Analisis Data," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, 2021.
- [8] E. Prasetyo and A. Nugroho, "Analisis Data Mining Menggunakan Metode CRISP-DM untuk Pengolahan Data," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, 2021.
- [9] W. Saputra and M. H. Ramadhan, "Penerapan Data Mining untuk Analisis Data Menggunakan Metode Regresi," *J. Inform. dan Sist. Inf.*, 2022.
- [10] A. Rianti *et al.*, "CRISP-DM: Metodologi Proyek Data Science," *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, pp. 107–114, 2023, [Online]. Available: <https://ojs.uadb.ac.id/index.php/Senatib/article/view/3015>
- [11] M. D. Irrawati and M. Mukaramah, "Implementasi Metode Regresi Linear Berganda untuk Mengatasi Pelanggaran Asumsi Klasik," *Stud. Akuntansi, Keuangan, dan Manaj.*, vol. 3, no. 2, pp. 83–94, 2024, doi: 10.35912/sakman.v3i2.2743.