

Prediksi Harga Handphone Berbasis Algoritma Supervised Learning

Afrizal Zein¹, Fordiana Ekawati²

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang
Jl. Raya Puspatek, No. 10, Buaran, Serpong, Tangerang Selatan, Banten, Indonesia

Email : dosen01495@unpam.ac.id, 02688@unpam.ac.id

Abstrak

Perkembangan teknologi ponsel yang pesat menyebabkan meningkatnya variasi spesifikasi dan harga di pasar, sehingga menimbulkan tantangan dalam menentukan harga yang sesuai untuk perangkat baru maupun bekas. Penelitian ini bertujuan untuk membangun model prediksi harga ponsel menggunakan algoritma supervised learning. Beberapa fitur utama yang digunakan meliputi spesifikasi teknis seperti kapasitas RAM, penyimpanan internal, ukuran layar, resolusi kamera, dan jenis prosesor. Metodologi yang diterapkan mencakup eksplorasi data, pra-pemrosesan, serta pelatihan dan evaluasi model menggunakan algoritma regresi linier, decision tree, dan random forest. Evaluasi performa dilakukan menggunakan metrik seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan R^2 Score. Percobaan ini menggunakan data yang diunduh dari Kaggle yang berisi 145 harga dan fitur ponsel. Ditemukan bahwa algoritma regresi linier dan hutan acak dapat memberikan prediksi yang relatif baik untuk ponsel dengan skor MAPE di bawah 10% dan skor R^2 di atas 95%. Metode hutan acak memprediksi harga sedikit lebih baik daripada regresi linier. Model random forest memberikan performa terbaik dalam memprediksi harga ponsel dengan tingkat akurasi yang tinggi. Temuan ini dapat dimanfaatkan oleh pelaku industri, penjual, maupun konsumen untuk menentukan harga pasar secara lebih objektif dan efisien.

Kata kunci: prediksi harga, supervised learning, ponsel, machine learning, random forest

Abstract

The rapid development of mobile phone technology has led to an increase in the variety of specifications and prices in the market, thus creating challenges in determining the appropriate price for new and used devices. This study aims to build a mobile phone price prediction model using a supervised learning algorithm. Some of the main features used include technical specifications such as RAM capacity, internal storage, screen size, camera resolution, and processor type. The methodology applied includes data exploration, pre-processing, and model training and evaluation using linear regression, decision tree, and random forest algorithms. Performance evaluation is carried out using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 Score. This experiment uses data downloaded from Kaggle containing 145 mobile phone prices and features. It was found that the linear regression and random forest algorithms can provide relatively good predictions for mobile phones with MAPE scores below 10% and R^2 scores above 95%. The random forest method predicts prices slightly better than linear regression. The random forest model provides the best performance in predicting mobile phone prices with a high level of accuracy. These findings can be utilized by industry players, sellers, and consumers to determine market prices more objectively and efficiently.

Keywords: price prediction, supervised learning, mobile phones, machine learning, random forest

1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi yang pesat telah mendorong pertumbuhan industri perangkat mobile, khususnya ponsel pintar (smartphone). Ponsel tidak lagi hanya berfungsi sebagai alat komunikasi, melainkan telah menjadi

perangkat multifungsi yang menunjang aktivitas sehari-hari, mulai dari hiburan, pekerjaan, hingga transaksi keuangan digital. Akibatnya, pasar ponsel menjadi sangat dinamis dengan variasi harga yang dipengaruhi oleh beragam spesifikasi teknis dan fitur.

Perkembangan teknologi yang pesat telah mendorong peningkatan permintaan terhadap perangkat komunikasi, khususnya ponsel pintar (smartphone). Di Indonesia, penggunaan smartphone terus meningkat, tidak hanya untuk komunikasi, tetapi juga untuk kegiatan ekonomi, pendidikan, dan hiburan. Namun, konsumen sering kali menghadapi kesulitan dalam menentukan pilihan yang tepat karena banyaknya variasi merek, spesifikasi, dan harga yang ditawarkan di pasaran. Kebingungan ini diperparah dengan kurangnya informasi yang objektif mengenai faktor-faktor yang memengaruhi harga sebuah ponsel. Menurut Zebua (2022), masyarakat Indonesia sering mengalami kebingungan ketika dihadapkan pada berbagai pilihan harga ponsel. Oleh karena itu, diperlukan suatu metode yang dapat membantu memprediksi harga ponsel secara lebih akurat. Salah satu pendekatan yang dapat digunakan adalah metode regresi linier, yang mampu menganalisis hubungan antara harga ponsel dengan variabel-variabel seperti spesifikasi teknis, merek, tahun rilis, dan fitur lainnya. Penelitian ini penting untuk memberikan kontribusi bagi dunia akademik dalam pengembangan model prediksi, bagi industri dalam strategi penetapan harga, serta bagi konsumen dalam mengambil keputusan pembelian yang lebih rasional. Salah satu tantangan dalam industri ini adalah menentukan harga jual ponsel yang kompetitif namun tetap menguntungkan. Penetapan harga yang tepat membutuhkan analisis terhadap berbagai faktor seperti kapasitas RAM, penyimpanan internal, ukuran layar, resolusi kamera, daya tahan baterai, hingga merek dan tren pasar saat ini. Dalam konteks ini, pemanfaatan teknologi kecerdasan buatan, khususnya pembelajaran mesin (machine learning), dapat memberikan solusi yang efisien dan akurat dalam memperkirakan harga ponsel berdasarkan data historis.

Supervised learning merupakan salah satu pendekatan dalam pembelajaran mesin yang bekerja dengan mempelajari hubungan antara data input (fitur) dan output (label) berdasarkan data pelatihan

yang telah dilabeli sebelumnya (Goodfellow et al., 2016).

Dalam kasus prediksi harga ponsel, supervised learning digunakan untuk membangun model regresi yang mampu memetakan spesifikasi teknis ponsel terhadap harga pasarnya. Algoritma seperti regresi linier, decision tree, random forest, hingga gradient boosting sering digunakan dalam studi-studi serupa untuk melakukan prediksi harga produk atau properti (Kuhn & Johnson, 2013).

Penelitian ini bertujuan untuk mengembangkan model prediksi harga ponsel berbasis algoritma supervised learning dengan memanfaatkan dataset spesifikasi ponsel dan harga aktualnya. Diharapkan model yang dihasilkan dapat membantu produsen, penjual, maupun konsumen dalam memperkirakan harga yang wajar untuk sebuah perangkat berdasarkan fitur-fiturnya.

2. Metodologi

Penelitian ini menggunakan Algoritma Supervised Learning dengan metode peramalan dengan memanfaatkan algoritma Linear Regression dan Random Forest Regression dengan dua skenario distribusi data yaitu 80:20 dan 70:30 untuk menentukan kinerja algoritma terbaik. Metode penelitian ini memiliki langkah-langkah terperinci seperti yang disajikan pada Gambar 1

Berikut adalah deskripsi langkah-langkah utama yang dapat digambarkan dalam Flowchart Proses Perancangan Penelitian Prediksi Harga Ponsel Berbasis Algoritma Supervised Learning



Gambar 1. Langkah-langkah Penelitian

a. Data Acquisition

Tahap akuisisi data merupakan langkah awal yang krusial dalam pengembangan model prediksi harga ponsel berbasis algoritma supervised learning. Data yang digunakan dalam penelitian ini diperoleh dari sumber terbuka (open-source) yang tersedia di internet, seperti Kaggle, situs e-commerce (misalnya Tokopedia, Shopee, dan Bukalapak), serta situs resmi produsen smartphone. Dataset yang dikumpulkan mencakup berbagai informasi spesifik terkait ponsel, antara lain:

1. Merek dan model ponsel
2. Ukuran RAM (dalam GB)
3. Kapasitas penyimpanan internal (ROM)
4. Ukuran layar
5. Resolusi kamera utama dan depan
6. Kapasitas baterai
7. Jenis prosesor
8. Sistem operasi
9. Tahun rilis
10. Harga ponsel

Data dikumpulkan dalam format CSV dan selanjutnya dilakukan proses pembersihan (data cleaning) untuk menghilangkan duplikasi, menangani nilai kosong (missing values), serta mengubah data kategorikal menjadi numerik menggunakan teknik encoding agar dapat digunakan oleh algoritma supervised learning. Setelah tahap preprocessing, data kemudian dibagi menjadi dua bagian: data latih (training data) dan data uji (testing data) dengan perbandingan umum 80:20.

b. Data selection

Seleksi data merupakan proses penting dalam memastikan bahwa hanya fitur-fitur (variabel) yang relevan dan signifikan yang digunakan dalam proses pelatihan model prediksi. Pada penelitian ini, setelah dilakukan proses akuisisi dan pembersihan data, dilakukan pemilihan atribut yang dianggap memiliki pengaruh langsung terhadap harga ponsel. Beberapa atribut yang dipilih untuk digunakan dalam model prediksi antara lain:

1. Brand (Merek Ponsel): Merek dapat memengaruhi persepsi kualitas dan harga.
2. RAM (Random Access Memory): Kapasitas RAM menentukan kinerja multitasking ponsel.
3. Storage (ROM/Internal Memory): Semakin besar kapasitas penyimpanan, biasanya semakin tinggi harga.
4. Screen Size (Ukuran Layar): Ukuran layar menjadi salah satu daya tarik utama konsumen.
5. Camera Quality (Resolusi Kamera Utama dan Kamera Depan): Kamera berkualitas tinggi sering kali meningkatkan harga.
6. Battery Capacity (Kapasitas Baterai): Baterai besar menjadi nilai tambah bagi pengguna.
7. Processor Type (Jenis Prosesor): Prosesor yang lebih canggih biasanya terdapat pada ponsel dengan harga lebih tinggi.

c. Data Cleaning

Data cleaning bertujuan untuk meningkatkan kualitas data agar model supervised learning (seperti regresi, decision tree, random forest, SVM, dll) bisa melakukan prediksi harga handphone secara akurat dan andal.

d. Data Transformation

Proses transformasi data adalah proses menggabungkan data ke dalam bentuk yang sesuai. Dalam proses ini, data numerik akan diubah menjadi rentang nilai 0-1 untuk semua atribut. Transformasi data digunakan untuk memproses atau menyesuaikan data sehingga kita dapat secara efektif menggunakan algoritma Regresi Linier dan juga Regresi Hutan Acak.

d. Implementation Linear Regression

Studi ini juga melakukan regresi multivariabel. Baterai, Frekuensi CPU, Inti CPU, Kamera Depan, Memori Internal, Kepadatan Piksel Ponsel (ppi), RAM, Kamera Belakang, resolusi, jumlah penjualan, ketebalan, dan berat diidentifikasi sebagai prediktor, sedangkan harga adalah responsnya.

$$\text{Harga} = a + bX$$

Harga = variabel respon

a = konstanta

b = koefisien regresi

X = variabel prediktor

```
%MULTIPLE VARIABLE REGRESSION
X = [battery, CPU_Freq, CPU_Core, Front_Camera,
    InternalMem, ppi, ram, RearCam, resolution, Sale,
    thickness, weight]

%Fit linear model regression
model = fitlm (X, Harga )
```

Untuk meningkatkan akurasi pada kumpulan data berdimensi rendah hingga sedang, kami menggunakan “fitlm” untuk menyesuaikan model regresi linier.

e. Evaluation Model

Evaluasi model dalam regresi berbeda dengan klasifikasi. Dalam pencarian model regresi, evaluasi model yang dicari adalah nilai error dalam suatu model. Evaluasi model regresi dilakukan dengan beberapa cara, antara lain: a. RMSE: mengevaluasi model dengan melihat tingkat error hasil prediksi. b. R2: mengevaluasi proporsi variabel dalam data yang dihitung untuk melihat kecocokan model. c. MAPE: mengevaluasi secara statistik keakuratan prediksi dalam metode peramalan, dengan membandingkan error ramalan dengan nilai sebenarnya.

3. Hasil dan Pembahasan

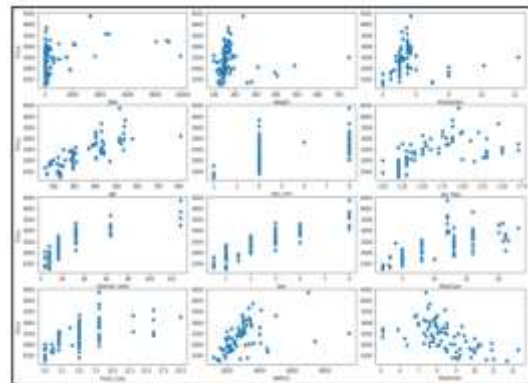
Analisis perancangan sistem memerlukan tahapan yang sistematis untuk mendapatkan prediksi yang baik dan bersesuaian dengan kegunaan dan tujuannya

3.1 Analisis Data

Sebelum memulai proses pemodelan, data mentah dengan variabel yang tidak diperlukan seperti Product_id dan sale dihilangkan sementara, dan rekaman dengan nilai yang hilang juga dihapus.

3.2 Visualisasi Data

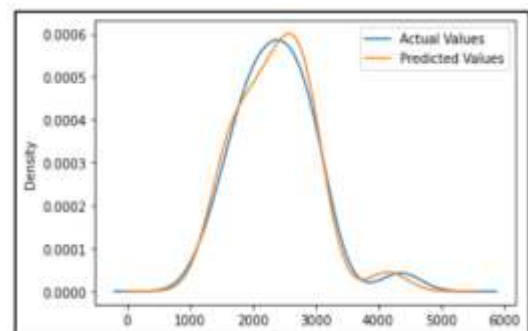
Tahap ini digunakan untuk melihat korelasi antara variabel Harga dengan variabel lainnya. Hasilnya berupa plot yang dapat dilihat pada Gambar 2.



Gambar 2. Korelasi Antar Variabel

3.3 Density Estimation

Pada tahap ini dilakukan distribusi probabilitas antara nilai aktual dan nilai prediksi variabel Harga. Hasilnya disajikan pada Gambar 3.



Gambar 3. Visualisasi Estimasi Kepadatan Kernel

3.4 Linear Regression

Tabel 1 menjelaskan bahwa resolusi dan berat tidak berkorelasi signifikan dengan Harga. Namun, yang paling berkorelasi dengan Harga adalah RAM, diikuti oleh memori internal, kepadatan piksel ponsel (ppi), kamera belakang, kamera depan, frekuensi CPU, ketebalan, inti CPU, dan baterai.

Tabel 1

Regresi Linier Dengan Satu Variabel Terhadap Harga

Variables	R2	P-Value
Battery	0.194	0.000
CPU Core	0.288	0.000
CPU Frequency	0.311	0.000
Front camera	0.395	0.000
Internal Memory	0.627	0.000
Phone Pixel Density (ppi)	0.590	0.000
RAM	0.785	0.000
Rear Camera	0.425	0.000
Resolution	0.008	0.289
Thickness	0.309	0.000
Weight	0.000	0.925

Regresi linier dengan beberapa variabel juga dilakukan. Hasil estimasi koefisien ditampilkan pada Tabel III. Variabel bebas (X) terdiri dari dua belas variabel beserta satu variabel terikat (Harga), yaitu: baterai, Frekuensi CPU, Inti CPU, Kamera Depan, Memori Internal, ppi, ram, Kamera Belakang, resolusi, ketebalan, dan berat. Selain itu, terbukti bahwa kamera depan, kamera belakang, dan berat tidak berkorelasi signifikan terhadap harga. Selain itu, ketika variabel resolusi dan ketebalan menurun, harga cenderung meningkat. Kami menemukan nilai R² lebih besar dari 0,9, dan nilai p jauh lebih kecil dari tingkat signifikansi default. Kita dapat menyimpulkan bahwa hubungan regresi linier yang signifikan memang ada antara variabel harga respons dan variabel prediktor dalam X.

Tabel 2.
Koefisien Estimasi Harga Telepon Seluler

Variables	Estimate	SE	tStat	pValue
(Intercept)	2197.200	266.440	8.247	0.000
Battery	0.145	0.031	4.675	0.000
CPU Frequency	121.250	48.582	2.496	0.014
CPU Core	51.170	10.346	4.946	0.000
Front Camera	7.137	5.109	1.397	0.165
Internal Memory	5.112	1.172	4.363	0.000
Phone Pixel Density (ppi)	1.081	0.225	4.814	0.000
RAM	96.088	26.395	3.640	0.000
Rear Camera	2.697	4.330	0.623	0.534
Resolution	-149.270	54.411	-2.743	0.007
Thickness	-94.354	14.868	-6.346	0.000
Weight	0.264	0.810	0.325	0.745

3.5 Model Evaluasi

Pemodelan pengujian pada penelitian ini menggunakan desain pemodelan partisi dengan data latih dan data uji sebesar 80:20 dan 70:30. Kemudian dilakukan pengujian prediksi dengan menggunakan Random Forest Regressor dengan hasil sebesar 95,53% sedangkan pengujian prediksi menggunakan Linear Regression dengan hasil sebesar 95,11% sebagaimana tercantum pada Tabel 3

Tabel 3.Perbandingan Model Akurasi

Method	Accuracy Model	
	80:20	70:30
Random Forest Regression	95.55%	94.29%
Linear Regression	95.11%	95.54%

Berdasarkan hasil penelitian yang tertera pada Tabel 4, diperoleh nilai akurasi yang telah diuji yaitu partisi dengan data training dan testing sebesar 80:20 dan 70:30.

Tabel 4.Perbandingan Evaluasi Model Regresi

Method	RMSE		MAPE		R ²	
	80:20	70:30	80:20	70:30	80:20	70:30
Random Forest Regression	148.73	179.30	6.10%	7.31%	95.53%	94.29%
Linear Regression	155.50	158.67	6.91%	6.58%	95.11%	95.54%

Berdasarkan hasil penelitian yang ditabulasikan pada Tabel 5, diperoleh nilai akurasi yang telah diuji adalah partisi dengan data training dan testing sebesar 80:20 dan 70:30. Aries Maesya, Yanfi Yanfi dan Lukas Hak Cipta @ Roman Science Publications Inc. Vol.5, No.1, Januari, 2023 Jurnal Internasional Teknik Terapan & Teknologi 45 Pada pengujian partisi data training sebesar 80% dan data testing 20%, nilai akurasi terbaik adalah Random Forest Regression sebesar 95,55%, sedangkan partisi data training 70% dan data testing 30 menghasilkan nilai akurasi terbaik yaitu Linear Regression sebesar 95,54%. Berdasarkan hasil penelitian yang diperoleh, dapat disimpulkan bahwa kedua metode prediksi yang selama ini digunakan untuk memprediksi harga ponsel yaitu Random Forest Regression dan Linear Regression mampu memberikan nilai error yang baik dengan nilai MAPE dibawah 10% dan nilai r² diatas 94%. Diantara kedua model tersebut, model Random Forest Regression mendapatkan hasil prediksi yang lebih baik dengan nilai R² sebesar 95,53%, nilai MAPE sebesar 6,1%, dan nilai RMSE sebesar 148.733. Karena pada regresi Random Forest terdapat parameter-parameter yang menentukan banyaknya keputusan yang akan digunakan.

4. Kesimpulan

Berdasarkan hasil penelitian yang diperoleh, dapat disimpulkan bahwa dari regresi linier variabel tunggal maupun variabel jamak, resolusi dan bobot tidak berhubungan secara signifikan terhadap Harga. Selain itu, kedua metode prediksi yang selama ini digunakan untuk memprediksi harga ponsel yaitu Random Forest Regression dan Linear Regression dapat memberikan nilai error yang baik dengan nilai MAPE dibawah 10% dan nilai R² diatas 94%. Diantara kedua model tersebut, model Random

Forest Regression mendapatkan hasil prediksi yang lebih baik dengan nilai R2 sebesar 95,53%, nilai MAPE sebesar 6,1%, dan nilai RMSE sebesar 148.733. Karena pada Random Forest Regression terdapat parameter yang menentukan banyaknya keputusan yang akan digunakan. Untuk penelitian selanjutnya, sebaiknya menambah sumber data tambahan dan mencoba metode supervised learning yang berbeda.

5. Daftar Pustaka

- A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, and G. Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique," in *Procedia Computer Science*, 2021, vol. 199. doi: 10.1016/j.procs.2022.01.100.
- Afrizal Zein, 2022. Evaluasi Keamanan Wireless LAN Menggunakan Issaf (Information System Security Assessment Framework), Sainstech: Jurnal Penelitian dan Pengkajian Sains dan Teknologi: Vol. 32 No. 2: DOI: <https://doi.org/10.37277/stch.v32i2>
- A. Gupta and T. J. Nagalakshmi, "Stock price prediction using linear regression in machine learning," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, 2019, doi: 10.35940/ijitee.L3932.1081219.
- A. Kaushal and A. Shankar, "House Price Prediction Using Multiple Linear Regression," *SSRN Electronic Journal*, 2021, doi: 10.2139/ssrn.3833734. Bankmycell.com, "How many smartphones are in the world?" <https://www.bankmycell.com/blog/how-many-phones-are-in-the-world> (accessed Oct. 09, 2022).
- "List of mobile phone brands by country," Wikipedia. https://en.wikipedia.org/wiki/List_of_mobile_phone_brands_by_country (accessed Oct. 09, 2022).
- E. Güvenç, G. Çetin, and H. Koçak, "Comparison of KNN and DNN Classifiers Performance in Predicting Mobile Phone Price Ranges," *Advances in Artificial Intelligence Research (AAIR)*, vol. 1, no. 1, 2021. I. M. Nasser and M. Al-Shawwa, "Developing Artificial Neural Network for Predicting Mobile Phone Price Range," *International Journal of Academic Information Systems Research*, vol. 3, no. 2, 2019.
- L. Breiman, "Random forests," *Mach Learn*, vol. 45, no. 1, 2001, doi: 10.1023/A:1010933404324.
- L. Rokach and O. Maimon, *Data mining with decision trees: Theory and applications*. 2007. doi: 10.1142/6604.
- Liu, J. Huang, H. Han, and H. Yang, "An Improved Intelligent Pricing Model for Recycled Mobile Phones," 2020. doi: 10.1109/CAC51589.2020.9327611.
- M. Cetin and Y. Koc, "Mobile Phone Price Class Prediction Using Different Classification Algorithms with Feature Selection and Parameter Optimization," 2021. doi: 10.1109/ISMSIT52890.2021.9604550.
- P. Sadowsky, "A Random Forests Approach to Predicting Clean Energy Stock Prices," *Journal of Risk and Financial Management*, vol. 14, no. 2, 2021, doi: 10.3390/jrfm14020048.
- Samsu Supriyatna, Salman Farizy, (2024) Perancangan dan Implementasi Aplikasi Monitoring Berkas Pencairan Dana Berbasis Web Menggunakan Metode Rapid Application Development, Sainstech: Jurnal Penelitian dan Pengkajian Sains Dan Teknologi: Vol. 34 No. 3: pp. DOI: <https://doi.org/10.37277/stch.v34i3.2078>.
- T. K. Ho, "Random decision forests," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 1995*, vol. 1. doi: 10.1109/ICDAR.1995.598994.
- T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans Pattern Anal Mach Intell*, vol. 20, no. 8, 1998, doi: 10.1109/34.709601.
- V. Khandelwal, A. K. Chaturvedi, and C. P. Gupta, "Amazon EC2 Spot Price Prediction Using Regression Random Forests," *IEEE Transactions on Cloud Computing*, vol. 8, no. 1, 2020, doi: 10.1109/TCC.2017.2780159.
- Q. Zhang, "Housing Price Prediction Based on Multiple Linear Regression," *Sci Program*,

Prediksi Harga Handphone Berbasis Algoritma Supervised Learning
Author : Afrizal Zein, Fordiana Ekawati- Sainstech Vol. 35 No. 2 (Juni 2025): 27-33
DOI: <https://doi.org/10.37277/stch.v35i2.2343>

vol. 2021, 2021, doi:
10.1155/2021/7678931.