

IMPLEMENTASI ALGORITMA RANDOM FOREST REGRESSION DALAM PREDIKSI HARGA LAPTOP

Siska Septianah¹, Ghita Athalina²

^{1,2} Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya
ghitaathalina.unsri@gmail.com

ABSTRAK

Penggunaan laptop telah menjadi kebutuhan signifikan dalam kehidupan modern, baik untuk keperluan pribadi, bisnis, maupun pendidikan. Pasar laptop terus berkembang pesat setiap tahunnya, didorong oleh kemajuan teknologi yang menghasilkan beragam jenis laptop dengan spesifikasi dan harga yang bervariasi. Prediksi harga laptop menjadi tantangan penting dalam analisis data, terutama dengan semakin beragamnya spesifikasi dan merek di pasar. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi harga laptop dan mengimplementasikan algoritma Random Forest Regression untuk melakukan prediksi harga secara akurat. Dataset yang digunakan mencakup berbagai atribut seperti RAM, CPU, GPU, resolusi layar, jenis laptop (TypeName), ukuran memori, dan berat laptop. Hasil analisis menunjukkan bahwa atribut RAM, CPU, GPU, dan resolusi layar memiliki pengaruh signifikan terhadap harga laptop, sementara TypeName dan ukuran layar (inches) tidak memberikan kontribusi yang berarti. Namun, ukuran memori dan berat laptop menjadi faktor penting yang dipertimbangkan. Model Random Forest Regression memberikan performa prediksi yang baik dengan nilai akurasi (R-squared) sebesar 0,83.

Keyword : *Random Forest, Prediksi Harga, Fitur Laptop*

1. PENDAHULUAN

Dalam era digital saat ini, penggunaan laptop telah menjadi kebutuhan yang signifikan baik untuk keperluan pribadi, bisnis, maupun pendidikan. Setiap tahunnya, pasar laptop terus berkembang dengan cepat seiring dengan kemajuan teknologi yang mendorong munculnya berbagai jenis laptop dengan spesifikasi dan harga yang bervariasi [1]. Bagi konsumen, memahami faktor-faktor yang mempengaruhi harga laptop dapat menjadi tantangan, terutama ketika mencoba menyesuaikan anggaran dengan kebutuhan spesifik mereka. Oleh karena itu, prediksi harga laptop yang akurat berdasarkan spesifikasi dan fitur-fitur utama sangat penting untuk membantu konsumen membuat keputusan yang tepat dan efektif.

Di sisi lain, bagi para produsen dan retailer, analisis dan prediksi harga laptop dapat memberikan keuntungan kompetitif dalam menentukan harga yang optimal sesuai dengan kondisi pasar. Prediksi merupakan sebuah proses memperkirakan secara sistematis sesuatu yang paling mungkin terjadi di masa yang akan datang berdasarkan kepada informasi yang dimiliki saat ini [2]. Prediksi bertujuan untuk mendapatkan jawaban terdekat terhadap sesuatu hal yang akan terjadi [3].

Penggunaan *machine learning*, khususnya algoritma berbasis pohon keputusan seperti Random Forest, mampu memberikan hasil prediksi yang akurat dengan proses pemilihan kumpulan data untuk meningkatkan kinerja klasifikasi sehingga dapat digunakan dalam berbagai bidang, termasuk analisis harga produk teknologi [4]. Random Forest merupakan metode ensemble yang memanfaatkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi dan mengurangi risiko *overfitting*

[5], sehingga algoritma ini sangat sesuai untuk masalah prediksi harga yang kompleks dengan banyak variabel independen seperti spesifikasi laptop. Model Random Forest mampu mengatasi kesalahan prediksi dalam decision tree dengan kebenaran yang didapatkan dari decision tree lainnya, karena pada model ini akan menggunakan decision tree yang tidak memiliki korelasi [6].

2. TINJAUAN PUSTAKA

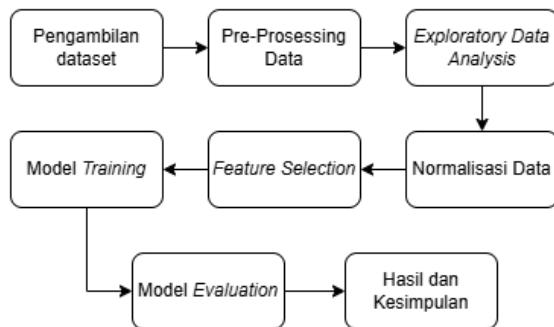
Di bidang e-commerce, akurasi prediksi harga laptop dapat membantu platform dalam menampilkan rekomendasi harga yang relevan bagi konsumen berdasarkan karakteristik spesifik dari laptop yang diminati. Beberapa penelitian terkait prediksi harga laptop telah dilakukan oleh penelitian sebelumnya, diantaranya oleh [7] yang melakukan analisis dan prediksi harga laptop berdasarkan variabel independennya menggunakan metode linear regression. Hasil akurasi prediksi yang didapatkan dalam penelitian tersebut sebesar 0.347. Selain itu penelitian yang dilakukan oleh [8], yang melakukan prediksi harga laptop dengan menggunakan model BiLSTM dan GRU yang diajukan untuk menyelesaikan pengelompokan yang telah dilakukan pada model regresi. Hasil dari penelitian tersebut yaitu, nilai koefisien determinasi pada model BiLSTM sebesar 0.205, lebih tinggi dibanding model GRU dengan nilai 0.181.

Berdasarkan pada beberapa penelitian diatas, maka penulis ingin melakukan analisis prediksi harga laptop dengan menggunakan algoritma Random Forest, penelitian ini diharapkan dapat mengembangkan model prediksi harga laptop yang robust dan efisien, yang tidak hanya berguna bagi konsumen tetapi juga bagi pelaku bisnis di industri

teknologi untuk mendukung pengambilan keputusan yang lebih terinformasi dan strategis.

3. METODE PENELITIAN

Dalam penelitian ini menggunakan metode Random Forest Regression yang terbagi dalam delapan tahapan penelitian yaitu pengambilan dataset, *pre-processing data*, *exploratory data analysis* (EDA), normalisasi data, *feature selection*, *training model*, *evaluation model*, dan hasil penelitian.



Gambar 1. Tahapan Penelitian

3.1. Pengumpulan Data

Penelitian ini menggunakan dataset “*Global Laptop selling dataset*” yang diunduh dari Kaggle <https://www.kaggle.com/datasets/mdriazhossain/global-laptop-selling-dataset>. Dataset ini mencakup 1303 data dengan 13 atribut. Atribut-atribut tersebut meliputi *Laptop_ID*, *Company*, *Product*, *TypeName*, *Inches*, *ScreenResolution*, *Cpu*, *Ram*, *Memory*, *Gpu*, *OpSys*, *Weight*, *Price_euros*. Dataset tersebut tersedia secara bebas untuk dapat digunakan, sesuai dengan lisensi yang ada yaitu *Open Database License* (OdbL) sehingga penggunaan dataset ini telah memenuhi persyaratan dan tidak memerlukan izin khusus tambahan untuk digunakan dalam penelitian.

3.2. Data Preprocessing

Tahap preprocessing data dilakukan untuk memastikan data yang akan digunakan dalam pemodelan telah sesuai dengan kebutuhan dan berkualitas. Tahapan ini mencakup proses penanganan *missing values*, *encoding* data kategorik, serta normalisasi data. *Missing values* dilakukan untuk menemukan nilai kosong yang dapat mempengaruhi akurasi pada model prediksi. *Encoding* data kategorik dilakukan dalam pemodelan untuk mengubah data kategorik menjadi numerik sehingga dapat di proses. Serta normalisasi data dilakukan untuk menyesuaikan fitur agar dapat mencegah dominasi antar fitur.

3.3. Exploratory Data Analysis

Exploratory Data Analysis (EDA) merupakan fase awal yang penting dalam memahami data yang tersedia. Dataset yang tersedia akan melewati tahap

analisis dasar yang dilakukan untuk mengetahui karakteristik maupun komponen penting dalam dataset sehingga dapat digunakan untuk membentuk model prediksi yang baik setelah analisis statistik. Tahapan ini bertujuan untuk memahami pola data dan hubungan antar variabel.

3.4. Feature Selection

Feature Selection merupakan tahapan untuk memilih fitur-fitur yang paling relevan dan informatif dalam dataset, yang memiliki pengaruh penting dalam model prediksi. Tujuan penggunaan tahapan ini adalah untuk mengurangi dimensi data serta meningkatkan kinerja model.

3.5. Evaluation Model

Dalam penelitian ini, akan dilakukan evaluasi terhadap model algoritma yang digunakan yaitu dengan melihat perbedaan nilai prediksi dan nilai asli pada dataset menggunakan *root mean square error* (RMSE) dan nilai koefisien korelasi. Melalui RMSE dapat dilihat besarnya penyimpangan yang terjadi antara nilai prediksi harga laptop dengan nilai actual harga laptop. Semakin tinggi nilai RMSE maka prediksi dinyatakan semakin baik [9], adapun RMSE dapat dihitung melalui rumus persamaan (1). Sedangkan nilai koefisien korelasi berguna untuk melihat besarnya hubungan antara total prediksi harga terhadap harga aktual.

$$RMSE = \sqrt{\frac{(x^1 - x)^2 + (y^1 - y)^2}{n}} \quad (1)$$

Dimana (x', y') merupakan nilai perhitungan, (x, y) merupakan nilai *exact*, dan n adalah jumlah data.

4. HASIL DAN PEMBAHASAN

Pada penelitian ini, peneliti melakukan pembangunan model prediksi dengan menggunakan *tools Google Collaboratory* dengan bahasa pemrograman *python*. Hasil dalam penelitian akan peneliti jelaskan pada bagian ini.

4.1. Cleaning Data

Library python yang digunakan dalam penelitian ini adalah *pandas*, *seaborn*, *matplotlib*, *math*, dan *numpy*. Adapun dataset yang digunakan terdiri atas 1303 baris data dan 13 atribut seperti terlihat pada Tabel 1.

Tabel 1. Deskripsi Atribut pada Dataset

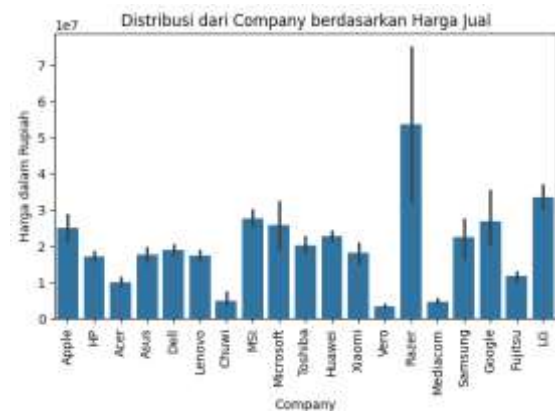
No.	Atribut	Deskripsi Data
1.	Laptop_ID	Nomor Urut
2.	Company	Perusahaan Laptop
3.	Product	Nama Produk Laptop
4.	TypeName	Jenis Laptop
5.	Inches	Ukuran Layar Laptop
6.	Screen Resolution	Jenis Layar Laptop
7.	Cpu	Jenis CPU yang digunakan
8.	RAM	Ukuran RAM pada Laptop

No.	Atribut	Deskripsi Data
9.	Memory	Jenis dan Ukuran Memory Laptop
10	Gpu	Jenis GPU yang digunakan
11	OpSys	Jenis Operation System pada Laptop
12	Weight	Ukuran Berat Laptop
13	Price_euros	Harga Laptop

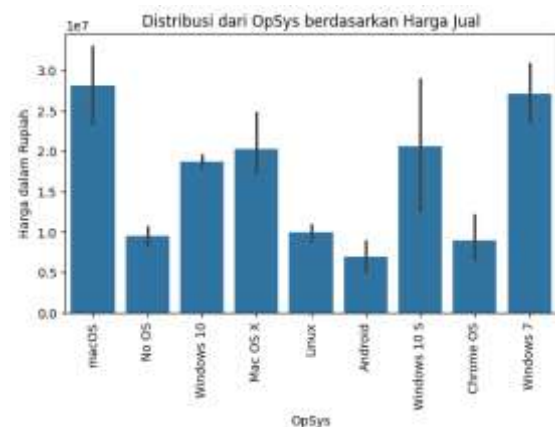
Dari dataset tersebut kemudian dilakukan pembersihan data dengan menghapuskan data duplikat serta data yang kurang diperlukan.

4.2. Exploratory Data Analysis

Pada bagian ini peneliti melakukan beberapa proses untuk memahami pengungkapan pada data, sebelum dilakukan model prediksi. Pengungkapan rata-rata harga laptop berdasarkan merek laptop ditunjukkan pada Gambar 2.



Gambar 2. Grafik Rata-rata Harga Laptop berdasar Perusahaan



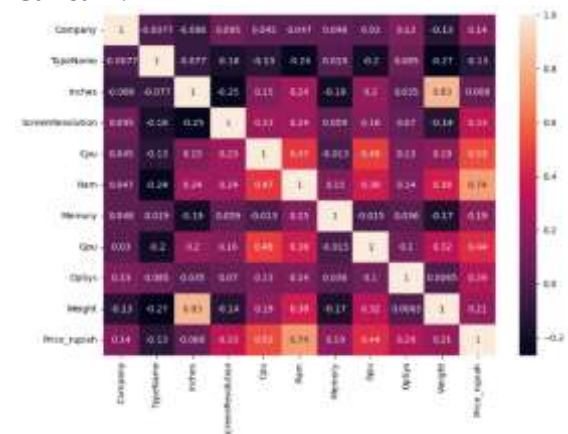
Gambar 3. Grafik Rata-rata Harga Laptop berdasar Sistem Operasi

Berdasarkan hasil yang ditampilkan, laptop dengan perusahaan razer menjadi salah satu laptop dengan harga yang rerata tinggi dibanding laptop lainnya. Sedangkan merek laptop dengan harga yang terjangkau adalah laptop produksi perusahaan vero. Berikutnya terdapat pengungkapan rata-rata harga laptop yang didasari oleh sistem operasi ditunjukkan

pada Gambar 3, dengan hasil kategori sistem operasi kelas atas sesuai dengan kebutuhan pengguna yaitu macOS. Kemudian sistem operasi Windows 7 yang masih menjadi daya tarik tinggi bagi kelas atas sesuai dengan kebutuhannya.

4.3. Feature Selection

Dalam tahapan ini, dilakukan pengelompokkan variabel-variabel independen yang memungkinkan memberikan pengaruh signifikan terhadap variabel dependen yaitu harga laptop, yang dilakukan menggunakan analisis korelasi dan divisualkan menggunakan heatmap untuk dapat menampilkan dengan jelas korelasi yang ada. Hasil analisis korelasi dengan visual heatmap dapat terlihat pada Gambar 4.



Gambar 4. Heatmap Korelasi Variabel

Hasil dari heatmap digunakan untuk melihat korelasi yang signifikan mempengaruhi harga laptop seperti variabel Ram dengan korelasi 0,74 dan variabel Cpu dengan korelasi 0,53. Sedangkan variabel yang tidak memiliki korelasi signifikan terhadap harga laptop yaitu variabel TypeName dengan korelasi sebesar -0,13.

4.4. Permodelan Dataset

Setelah melakukan cleaning dan pre-processing data, tahapan selanjutnya adalah pemodelan data yang dikelompokkan menjadi dua bagian yaitu data training dan data testing dengan perbandingan 90% : 10%. Sehingga didapatkan 1172 data yang digunakan dalam data training dan 131 data untuk data testing seperti terlihat pada Gambar 5.

```
train, test = train_test_split(df2, test_size=0.1, random_state=0)

print("No of training examples: {train.shape[0]}")
print("No of testing examples: {test.shape[0]}")

No of training examples: 1172
No of testing examples: 131
```

Gambar 5. Pemisahan Data Training dan Data Testing

Setelah melakukan pembagian data, dilanjutkan dengan penerapan algoritma random forest regression dalam permodelan data training

yang telah diperoleh dan dilanjutkan dengan uji akurasi model prediksi, dengan hasil dari pengujian seperti pada Tabel 2.

Tabel 2. Nilai Korelasi Model

Matrik	Nilai
R2	0.944

4.5. Hasil Prediksi

Hasil prediksi yang dihasilkan oleh model *random forest regression* dapat dilihat pada Gambar 6. Hasil prediksi menunjukkan bahwa beberapa prediksi sangat akurat dan sesuai dengan harga aktual.



Gambar 6. Grafik Hasil Prediksi

Tabel 3. Nilai Hasil Performa Model

Matrik	Nilai
R2	0.839
Mean Absolute Error (MAE)	3424103.0
Mean Squared Error (MSE)	263178403
Root Mean Squared Error (RMSE)	5130091.6

Dari hasil pengujian model prediksi pada Tabel 3 tersebut, berdasarkan data testing didapati nilai RMSE sebesar 513, nilai MAE sebesar 324, dan nilai MSE sebesar 263. Serta nilai tingkat akurasi prediksi dari algoritma *random forest regression* yang mencapai nilai 0.83 menunjukkan bahwa model ini dapat diandalkan untuk melakukan prediksi harga..

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dan analisis penelitian yang telah dilakukan, faktor yang mempengaruhi harga laptop adalah ukuran Ram, Cpu, Gpu, dan Screen Resolution. Hasil tersebut menunjukkan bahwa TypeName dan ukuran Inches laptop tidak signifikan menjadi hal yang mempengaruhi harga laptop. Namun ukuran Memory dan berat laptop menjadi hal yang dipertimbangkan. Dalam pemilihan algoritma *random forest regression* menunjukkan hasil prediksi dengan nilai akurasi sebesar 0.83 lebih baik dibanding penelitian sebelumnya [10] menggunakan algoritma regresi linear berganda yang menunjukkan nilai akurasi sebesar 0,68.

DAFTAR PUSTAKA

- [1] F. V. P. Samosir and H. Hery, "Analysis of Laptop Purchase History to Determine Key Factors Using Logistic Regression," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 2, Aug. 2024, doi: 10.28932/jutisi.v10i2.8897.
- [2] V. Arinal and M. Azhari, "Penerapan Regresi Linear Untuk Prediksi Harga Beras Di Indonesia," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 341–346, 2023, doi: 10.55338/saintek.v5i1.1417.
- [3] M. Yusuf, A. Setyanto, and K. Aryasa, "Analisis Prediksi Curah Hujan Bulanan Wilayah Kota Sorong Menggunakan Metode Multiple Regression," 2022.
- [4] Arifin Yusuf Permana, Hari Noer Fazri, M.Fakhrizal Nur Athoilah, Mohammad Robi, and Ricky Firmansyah, "Penerapan Data Mining Dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Jurnal Ilmiah Teknik Informatika dan Komunikasi*, vol. 3, no. 2, pp. 27–41, Jun. 2023, doi: 10.55606/juitik.v3i2.472.
- [5] E. Fitri, "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)*, vol. 4, no. 1, pp. 2723–1453, 2023, doi: 10.52158/jacost.491.
- [6] J. Sanjaya, E. Renata, V. E. Budiman, F. Anderson, and M. Ayub, "Prediksi Kelalaian Pinjaman Bank Menggunakan Random Forest dan Adaptive Boosting," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 1, Apr. 2020, doi: 10.28932/jutisi.v6i1.2313.
- [7] S. Nur Wardani and Nurmalitasari, "Penerapan Regresi Linier Berganda Untuk Memprediksi Harga Laptop Dengan Menggunakan Software Python," in *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, 2023, pp. 377–382.
- [8] E. Novela Waroi, A. Setyanto, and Khusnawi, "Prediksi Harga Laptop Menggunakan Algoritma GRU dan BILSTM," *Jurnal Sosial dan Teknologi (SOSTECH)*, vol. 4, no. 7, pp. 408–424, 2024.
- [9] Moch Farryz Rizkilloh and Sri Widiyanesti, "Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, pp. 25–31, Feb. 2022, doi: 10.29207/resti.v6i1.3630.
- [10] N. N. Sari, T. T. Anisah, and R. Fitriani, "Implementasi Machine Learning Untuk Prediksi Harga Laptop Menggunakan Algoritma Regresi Linear Berganda," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 14, no. 2, pp. 162–177, Jul. 2024, doi: 10.34010/jamika.v14i2.12923