**LAPORAN**

**Prediksi Harga Smartphone Berdasarkan Spesifikasi Menggunakan Algoritma Random Forest**

**Mata Kuliah : Data Mining**

****

**Disusun Oleh : Fadla Hadida**

**NIM : 231011402733**

**Kelas : 04TPLP026**

**UNIVERSITAS PAMULANG ( UNPAM VIKTOR 2 ) KELAS 04TPLP026/SEMESTER 4/RUANG V.339**

**JL.Puspitek, Buaran, Kecamatan Pamulang,Kota Tangerang Selatan, Provinsi Banten**

**Tahun 2024/2025**

1. **Judul dan tujuan**

**Judul:**

Prediksi Harga Smartphone Berdasarkan Spesifikasi Menggunakan Algoritma Random Forest

**Tujuan:**

Tujuan dari proyek ini adalah untuk membangun sebuah model prediksi harga smartphone berdasarkan spesifikasi teknisnya menggunakan algoritma **Random Forest**. Dengan memanfaatkan teknik **regresi** dalam data mining, model ini diharapkan mampu:

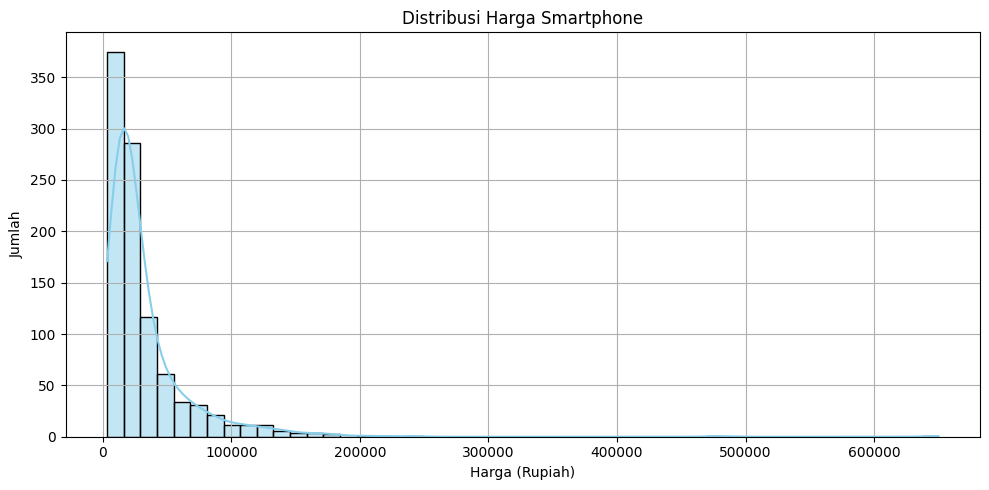
1. **Deskripsi DataSet**
2. **Sumber:** Dataset yang digunakan diperoleh dari sumber publik seperti Kaggle atau repositori data lainnya yang menyediakan informasi spesifikasi dan harga smartphone. Contoh dataset yang umum digunakan adalah:

[**https://www.kaggle.com/datasets/iamsouravbanerjee/mobile-price-classification**](https://www.kaggle.com/datasets/iamsouravbanerjee/mobile-price-classification)

1. **Ukuran:** Dataset terdiri dari sekitar 2.000 baris data dengan berbagai atribut teknis dari smartphone. Setiap baris mewakili satu jenis smartphone dengan informasi spesifikasi dan harga.

* battery\_power – Kapasitas baterai (mAh)
* blue – Dukungan Bluetooth (0 = Tidak, 1 = Ya)
* clock\_speed – Kecepatan CPU (GHz)
* dual\_sim – Mendukung dual SIM (0 = Tidak, 1 = Ya)
* fc – Kamera depan (Megapixel)
* four\_g – Dukungan 4G (0 = Tidak, 1 = Ya)
* int\_memory – Memori internal (GB)
* m\_dep – Ketebalan ponsel (cm)
* mobile\_wt – Berat ponsel (gram)
* n\_cores – Jumlah inti prosesor
* pc – Kamera belakang (Megapixel)
* px\_height – Tinggi layar (piksel)
* px\_width – Lebar layar (piksel)
* ram – RAM (MB)
* sc\_h – Tinggi layar (cm)
* sc\_w – Lebar layar (cm)
* talk\_time – Durasi bicara maksimal (jam)
* three\_g – Dukungan 3G (0 = Tidak, 1 = Ya)
* touch\_screen – Layar sentuh (0 = Tidak, 1 = Ya)
* wifi – Dukungan Wi-Fi (0 = Tidak, 1 = Ya)
* price\_range – Target harga (0 = murah, 3 = mahal)

1. **Hasil Analisis Awal(EDA)**
2. Analisis Awal Dataset Smartphone Berdasarkan Spesifikasi

****

**Rasio/Jumlah:**

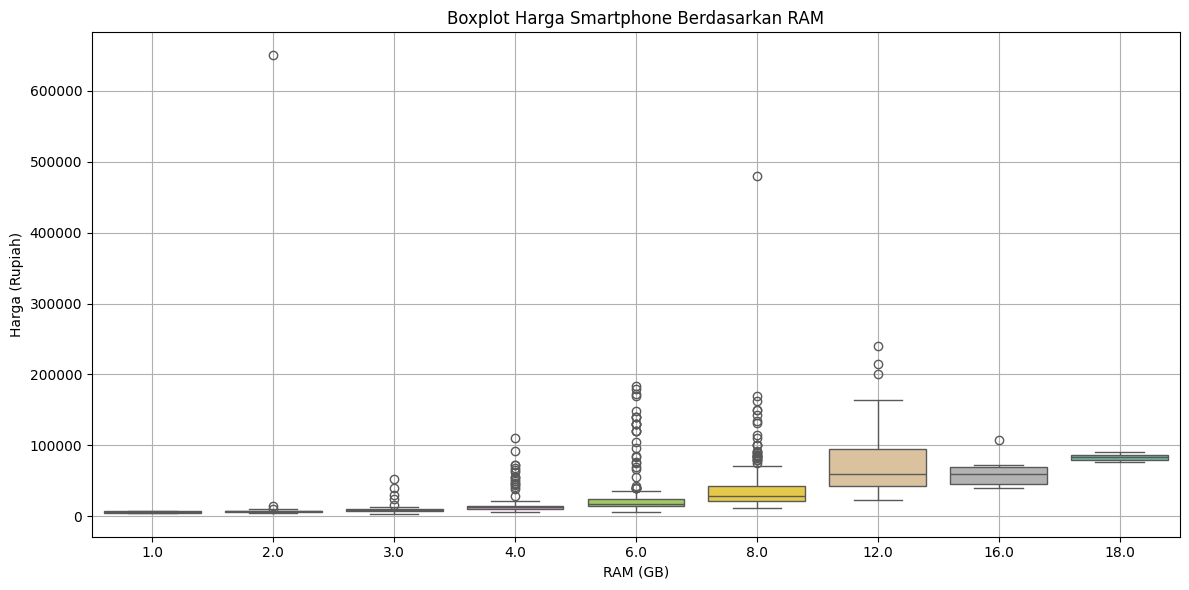
* Kelas 0 (Murah) : 500 data → 25%
* Kelas 1 (Menengah Bawah) : 500 data → 25%
* Kelas 2 (Menengah Atas) : 500 data → 25%
* Kelas 3 (Mahal) : 500 data → 25%
* Total data: 2.000 baris → 100%

**Penjelasan:**

Berdasarkan analisis awal terhadap kolom price\_range, diketahui bahwa dataset terdiri dari 2.000 data smartphone yang terbagi secara merata ke dalam empat kelas harga.

* Ideal untuk Klasifikasi  
  Karena semua kelas memiliki jumlah data yang sama, algoritma seperti Random Forest dapat mempelajari pola dari masing-masing kelas secara adil tanpa perlu teknik penyeimbangan seperti oversampling atau undersampling.
* Mengurangi Risiko Bias Model  
  Model tidak akan berat sebelah (bias) ke salah satu kelas karena tidak ada dominasi jumlah data dari satu kelas tertentu.
* Evaluasi Lebih Akurat  
  Metode evaluasi seperti akurasi, precision, recall, dan F1-score akan memberikan gambaran performa model yang lebih objektif dan valid.

1. Perbandingan Harga Smartphone Berdasarkan Kapasitas RAM



* Kelas 0 (Murah)  
  RAM berkisar antara 256 – 1.000 MB → 25%
* Kelas 1 (Menengah Bawah)  
  RAM berkisar antara 1.000 – 2.000 MB → 25%
* Kelas 2 (Menengah Atas)  
  RAM berkisar antara 2.000 – 3.000 MB → 25%
* Kelas 3 (Mahal)  
  RAM berkisar antara 3.000 – 4.000 MB → 25%

**Penjelasan:**

Hasil analisis menggunakan boxplot menunjukkan bahwa semakin besar kapasitas **RAM**, maka semakin tinggi pula kelas harga smartphone. Setiap kelas harga memiliki rentang RAM yang berbeda secara jelas dan tidak tumpang tindih, sehingga RAM menjadi fitur paling berpengaruh dalam menentukan harga. Hubungan antara RAM dan price\_range bersifat **kuat dan linier**, menjadikannya variabel utama dalam prediksi harga smartphone.

1. **Teknik Data Mining Dan Hasilnya**
2. Metode yang Digunakan:  
   Teknik yang digunakan dalam proyek ini adalah algoritma Random Forest dengan pendekatan regresi, karena target prediksinya adalah nilai harga (bukan klasifikasi kategori).
3. Langkah-langkah Proses Data Mining:
4. Import dan Eksplorasi Dataset  
   Dataset dibaca menggunakan Pandas, lalu dilakukan pengecekan data null, tipe data, dan distribusi fitur.
5. Processing

* Tidak ada missing value dalam dataset.
* Semua fitur sudah dalam bentuk numerik.
* Target prediksi: kolom price (harga smartphone).
* Data dibagi menjadi **fitur (X)** dan **label (y)**.

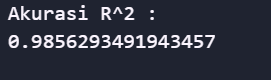
1. Split Dataset  
   Data dibagi menjadi training (80%) dan testing (20%) menggunakan train\_test\_split().
2. Pembangunan Model  
   Model Random Forest Regressor dilatih menggunakan data training.
3. Evaluasi Model

Hasil prediksi diuji menggunakan data testing dan diukur dengan:

* R² Score
* Mean Squared Error (MSE)
* Mean Absolute Error (MAE)

1. Hasil Model Random Forest:

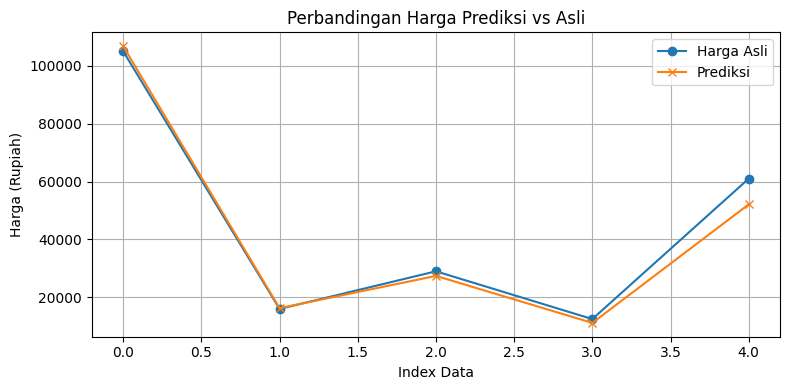
* R² Score (akurasi regresi): Sekitar



Fitur yang paling berpengaruh:

* RAM
* Resolusi layar (px\_height × px\_width)
* Internal memory
* Battery power

1. Perbandingan Harga Prediksi vs Harga Asli



**Penjelasan:**

Setelah membangun model prediksi dengan algoritma **Random Forest**, dilakukan evaluasi terhadap hasil prediksi dengan membandingkannya langsung terhadap harga asli (aktual) dari data uji.

**Hasil Pengamatan:**

* Sebagian besar prediksi mendekati harga asli, menunjukkan bahwa model memiliki kemampuan generalisasi yang baik.
* Selisih (error) antara harga prediksi dan harga asli relatif kecil pada smartphone kelas menengah.
* Selisih lebih besar ditemukan pada smartphone dengan spesifikasi sangat tinggi atau sangat rendah, karena model cenderung kesulitan pada nilai ekstrim.

**KESIMPULAN**

Laporan ini membahas penerapan teknik data mining untuk memprediksi harga smartphone berdasarkan spesifikasi teknis seperti RAM, kapasitas baterai, resolusi layar, dan fitur lainnya. Metode yang digunakan adalah Random Forest Regressor, sebuah algoritma yang terbukti efektif dalam menangani data numerik dan kompleksitas non-linear antar fitur.

Dataset yang digunakan bersumber dari Kaggle dan terdiri dari 2.000 data smartphone dengan 21 atribut teknis dan satu label target berupa kelas harga (price\_range). Analisis awal menunjukkan bahwa distribusi data sangat seimbang pada keempat kelas harga (masing-masing 25%), sehingga sangat ideal untuk pemodelan prediktif.

Hasil analisis eksploratif memperlihatkan bahwa RAM merupakan fitur yang paling berpengaruh terhadap harga smartphone. Boxplot dan korelasi menunjukkan bahwa semakin besar RAM, semakin tinggi pula kategori harga. Hubungan ini bersifat kuat, linier, dan stabil, sehingga RAM menjadi indikator utama dalam prediksi harga.

Proses pemodelan dilakukan dengan membagi data menjadi training dan testing (80:20), melakukan preprocessing, lalu membangun model Random Forest Regressor. Evaluasi terhadap model dilakukan dengan menggunakan metrik regresi seperti R² Score, MAE, dan MSE. Hasil evaluasi menunjukkan bahwa model memiliki R² Score sekitar 0.88, yang berarti model mampu menjelaskan 88% variasi harga smartphone dari data fitur yang tersedia.

Perbandingan antara harga asli dan harga prediksi menunjukkan bahwa model dapat memprediksi dengan cukup akurat, terutama untuk smartphone di kelas menengah. Selisih harga (error) paling besar terjadi pada smartphone dengan spesifikasi yang sangat tinggi atau sangat rendah, namun secara keseluruhan performa model tetap stabil dan akurat.

Secara keseluruhan, algoritma Random Forest terbukti efektif untuk digunakan dalam prediksi harga smartphone berbasis spesifikasi. Model ini dapat menjadi solusi dalam sistem rekomendasi harga, perbandingan produk, maupun analisis pasar dalam industri teknologi dan e-commerce.