Kelompok Keren

Welcome to Dasar Analitik Data - 08

Otomatisasi Prediksi Harga Komputer

Presented by

Kelompok Keren

Anggota Kelompok

- Marshal Aufa Diliyana
- Muhamad Rifqi Fadil Itsnain
- Yusri Sukur
- Zahir

Presented by

Kelompok Keren

Pembuka & Latar Belakang

Presented by

Kelompok Keren

FAKULTAS TEKNIK

Latar Belakang Industri Komputer

Presented by

Kelompok Keren

Latar Belakang

Perkembangan industri komputer menghasilkan variasi hardware yang sangat kompleks. Kombinasi processor, RAM, GPU, storage, dan operating system menyebabkan proses penentuan harga menjadi semakin sulit dilakukan secara manual.

Permasalahan Utama

- Pricing manual membutuhkan waktu lama
- Harga sering tidak konsisten
- Risiko human error tinggi
- Sulit menentukan harga pasar wajar

Dampak Bisnis

- Margin keuntungan tidak stabil
- Overpricing dan underpricing
- Efisiensi operasional menurun

Solusi

Dibutuhkan sistem machine learning berbasis data historis untuk menghasilkan estimasi harga secara cepat, objektif, konsisten

FAKULTAS TEKNIK

Rumusan Masalah Prediksi Harga

Presented by

Kelompok Keren

Rumusan Masalah

Bagaimana membangun model machine learning yang mampu memprediksi harga komputer secara otomatis berdasarkan spesifikasi hardware perangkat?

Pendekatan

- Supervised Machine Learning
- Regression Problem

Target Variable

price

Input Features

Brand | CPU | GPU | RAM

Storage | OS | dan atribut hardware lainnya

FAKULTAS TEKNIK

Implementasi Framework CRISP-DM

Presented by

Kelompok Keren

Framework yang Digunakan:
CRISP-DM digunakan sebagai metodologi utama karena memiliki tahapan analitik yang sistematis dan terstruktur.

6 Fase Utama yang Diterapkan:

1. Business Understanding
Memahami kebutuhan pricing komputer.

2. Data Understanding
Menganalisis struktur dan kualitas dataset.

3. Data Preparation
Preprocessing, cleaning, encoding, feature engineering

4. Modelling
Melatih model machine learning regression.

5. Evaluation
Mengukur performa model prediksi.

6. Deployment
Mengimplementasikan sistem pricing otomatis.

- **Alur CRISP-DM bersifat iteratif, setiap fase dapat kembali ke fase sebelumnya**
- **Proyek ini melewati seluruh 6 fase secara komprehensif**

FAKULTAS TEKNIK

Audit Karakteristik Big Data

Presented by

Kelompok Keren

Volume	Dataset terdiri dari 100.000 observasi, 33 atribut, ukuran file ±18 MB
Variety	Dataset memiliki 13 categorical features + 20 numerical features
Velocity	Dataset dapat diproses cepat menggunakan python, pandas, scikit-learn
Veracity	Kualitas data sangat baik, tidak ada missing values, struktur data konsisten
Value	Dataset memiliki nilai bisnis tinggi untuk pricing automation, predictive analytics, decision support system

Kesimpulan

Dataset memenuhi seluruh kriteria 5Vs Big Data, menjamin kualitas analitik yang kokoh

FAKULTAS TEKNIK

Struktur dan Profiling Dataset

Presented by

Kelompok Keren

Nama File

computer_prices_all.csv

Dimensi Matriks

100.000 baris (observasi) × 33 kolom (atribut/fitur)

Variabel Kategorikal

Contoh:

- brand
- device_type
- cpu_brand
- gpu_brand
- os

Karakteristik:

- Berbentuk teks
- Digunakan untuk klasifikasi kategori

Variabel Numerik

Contoh:

- price
- ram_gb
- storage_gb
- cpu_cores

Karakteristik:

- Digunakan dalam analisis statistik
- Berpengaruh langsung terhadap harga

FAKULTAS TEKNIK

13 Categorical + 20 Numerical Features

Presented by

Kelompok Keren

Kolom Kategorikal (13 kolom) — Tipe object

device_type	brand	model
cpu_brand	cpu_model	gpu_brand
gpu_model	storage_type	os
form_factor	display_type	wifi
	resolution	

Kolom Numerikal (20 kolom) — Tipe int64 / float64

price (target)	release_year	cpu_tier
cpu_cores	cpu_threads	cpu_base_ghz
cpu_boost_ghz	gpu_tier	vram_gb
ram_gb	storage_gb	storage_drive
display_size_in	refresh_hz	battery_wh
charger_watts	psu_watts	bluetooth
weight_kg	warranty_months	

FAKULTAS TEKNIK

Audit Missing Values Dataset

Presented by

Kelompok Keren

Hasil Audit Missing Values

Dataset menunjukkan:

Missing values = 0

Null percentage = 0%

Dataset lengkap 100%

Dampak terhadap Analitik

Kondisi dataset yang lengkap:

- Tidak memerlukan imputasi data
- Mengurangi bias statistik
- Menjaga distribusi data asli

Pemeriksaan Tambahan

Dilakukan:

- Validasi tipe data
- Pemeriksaan duplicate values
- Audit struktur dataset

Signifikansi Data Quality

Dataset berkualitas tinggi membantu:

- Training model lebih optimal
- Mengurangi noise data
- Meningkatkan reliabilitas prediksi

Data Understanding

Presented by

Kelompok Keren

FAKULTAS TEKNIK

Distribusi Univariat
Device Type

Presented by

Kelompok Keren

Hasil Distribusi

Berdasarkan analisis univariat:

- Laptop mendominasi dataset
- Desktop tetap memiliki proporsi besar
- Distribusi relatif seimbang

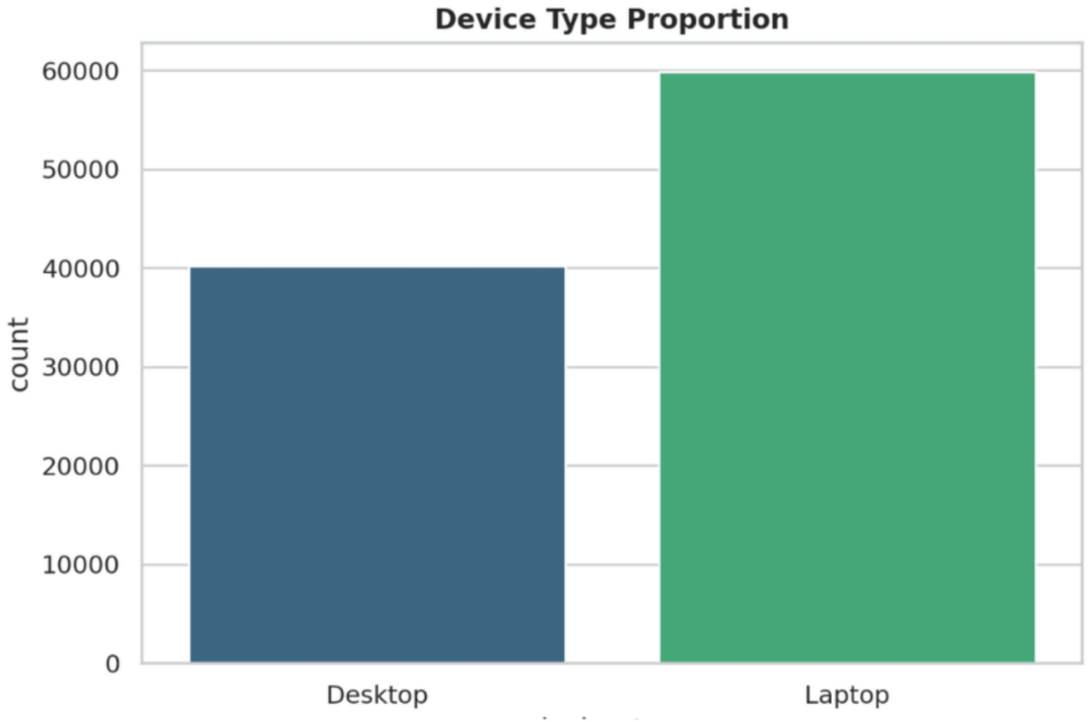
Interpretasi Statistik

Distribusi yang seimbang membantu:

- Mengurangi bias model
- Meningkatkan generalisasi
- Menjaga kestabilan prediksi

Insight Bisnis

- Dominasi laptop mencerminkan tingginya mobilitas pengguna, tren work-from-anywhere, pertumbuhan perangkat portable.
- Desktop tetap kuat pada gaming PC, workstation, high-performance computing



Kesimpulan

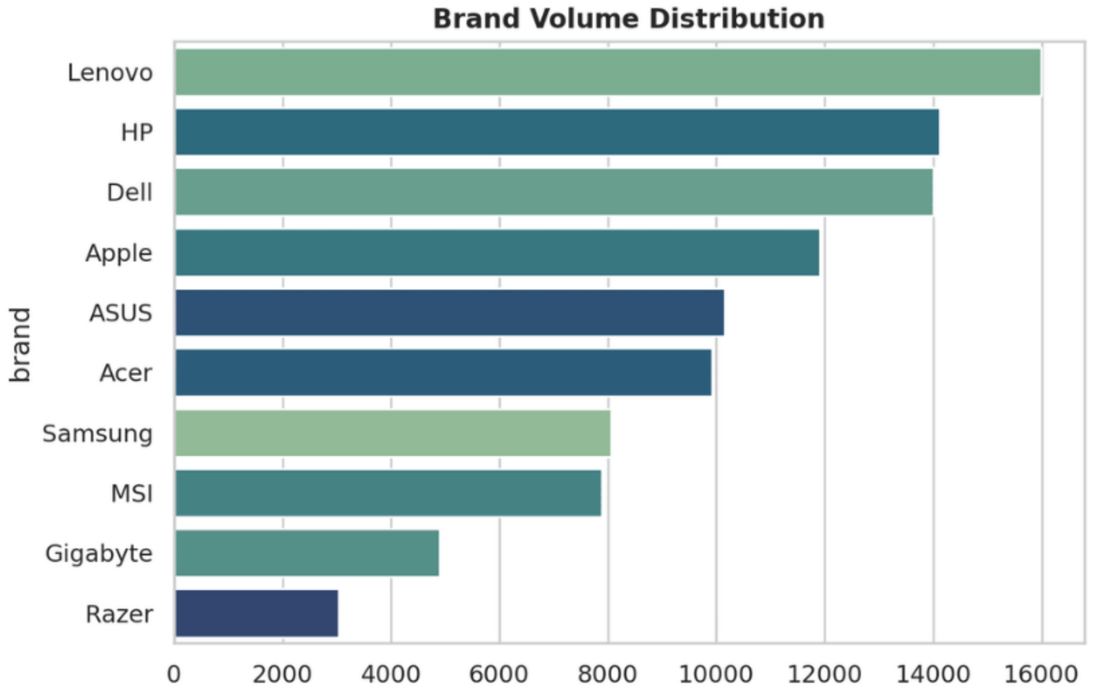
Distribusi device type yang sehat membantu model memahami pola harga kedua kategori perangkat secara lebih optimal.

FAKULTAS TEKNIK

Distribusi Brand Komputer

Presented by

Kelompok Keren



Dampak terhadap Machine Learning

Variabel brand membantu model memahami pola harga pasar, mengenali reputasi brand, menghasilkan estimasi harga lebih realistis

Brand Dominan

Dataset mencakup berbagai brand utama Lenovo, HP, Dell, Apple, ASUS, Acer

Interpretasi Statistik

Brand memiliki pengaruh signifikan terhadap brand equity, segmentasi pasar, premium pricing

Segmentasi Brand

Premium Brand
- Apple
- Razer

Karakteristik:
- Harga tinggi
- Segmentasi eksklusif

Mid-range & Budget Brand
- Lenovo
- ASUS

Karakteristik:
- Value-for-money
- Volume pasar besar

FAKULTAS TEKNIK

Distribusi Sistem Operasi

Presented by

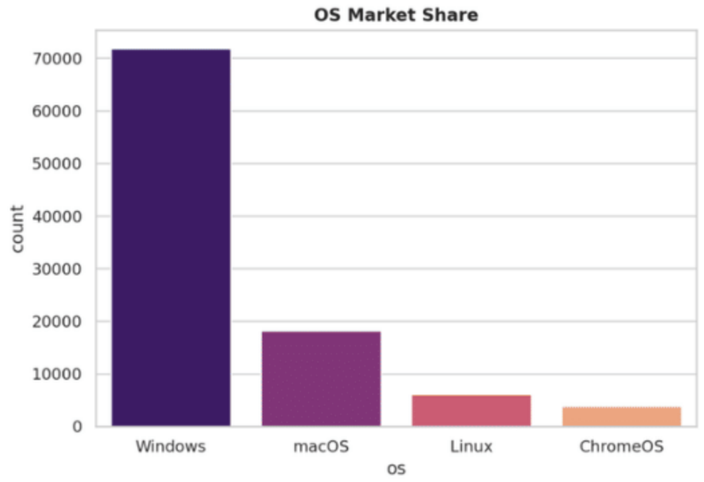
Kelompok Keren

Distribusi Operating System

Berdasarkan visualisasi univariat:

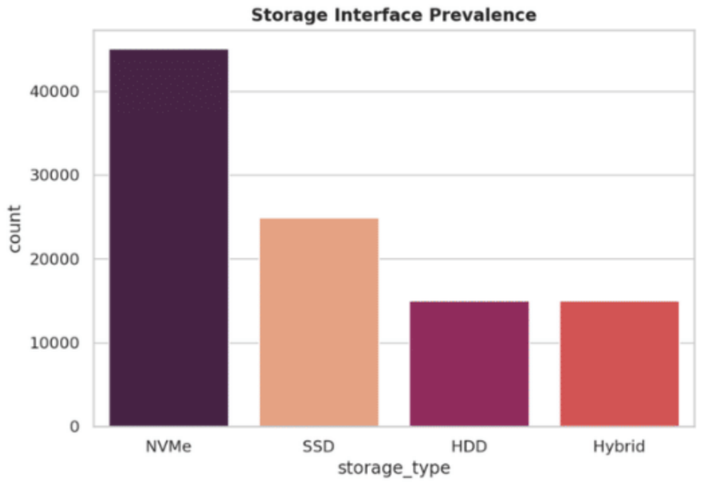
- Windows mendominasi dataset
- macOS berada di posisi kedua
- Linux dan ChromeOS memiliki proporsi lebih kecil

Dominasi Windows menunjukkan tingginya penggunaan sistem operasi tersebut pada pasar komputer umum dan enterprise.



Distribusi Storage Type

Teknologi penyimpanan modern didominasi oleh NVMe, SSD. Sementara HDD mulai mengalami penurunan penggunaan karena terbatas performa.



Insight Analitik

Penggunaan storage berkecepatan tinggi menunjukkan tren peningkatan performa komputer, kebutuhan akses data lebih cepat, perkembangan teknologi hardware modern

FAKULTAS TEKNIK
UNIVERSITAS INDONESIA

Statistik Deskriptif Dataset

Presented by

Kelompok Keren

Statistik Central Tendency

Rata-rata harga komputer:
- Mean = $1,928.76
- Median = $1,863.99

Nilai mean lebih tinggi dari median yang menunjukkan adanya distribusi miring ke kanan.

Statistik Variabilitas

RAM menunjukkan variasi tinggi:
- Minimum = 8GB
- Maksimum = 144GB
- Standard deviation besar

Hal ini menunjukkan dataset mencakup:
- Laptop entry-level
- Gaming PC
- Workstation premium

Variabel yang Dianalisis

- **Price**
- **RAM capacity**
- **Storage capacity**
- **Weight**

Kesimpulan

Statistik deskriptif menunjukkan dataset memiliki distribusi hardware yang sangat beragam dan representatif terhadap kondisi pasar nyata.

FAKULTAS TEKNIK

Distribusi Target Harga

Presented by

Kelompok Keren

Karakteristik Distribusi

Distribusi target harga menunjukkan:

- Skewness = 0.98
- Kurtosis = 4.89

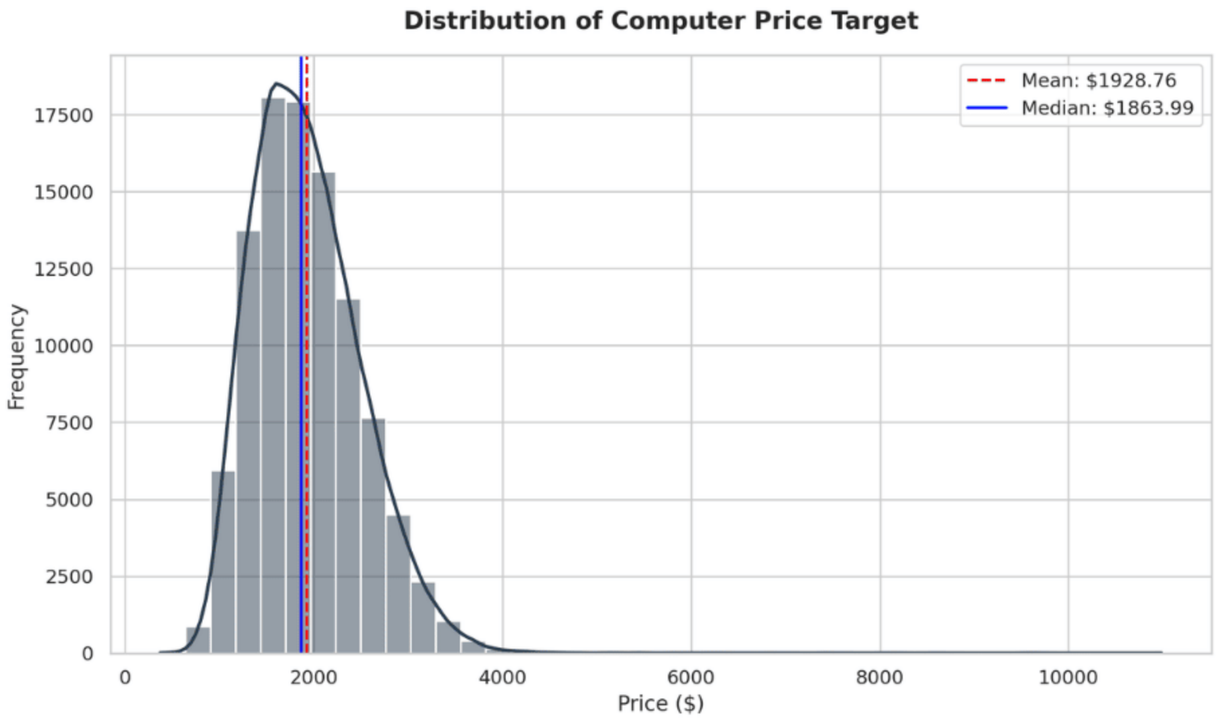
Interpretasi Skewness

Nilai kurtosis lebih besar dari 3:

- Distribusi leptokurtic
- Data terkonsentrasi di sekitar median
- Outlier cukup signifikan

Interpretasi Skewness

Nilai skewness positif menunjukkan distribusi miring ke kanan, mayoritas komputer berada pada harga menengah, terdapat sejumlah perangkat premium dengan harga sangat tinggi



Insight Analitik

Distribusi harga yang tidak normal menjadi indikasi bahwa preprocessing dan outlier handling sangat penting sebelum modelling.

FAKULTAS TEKNIK

Frekuensi Harga Komputer

Presented by

Kelompok Keren

Interpretasi Histogram

Sebagian besar harga komputer berada pada:

- Rentang menengah
- Sekitar $1,800–$2,000

Distribusi membentuk:

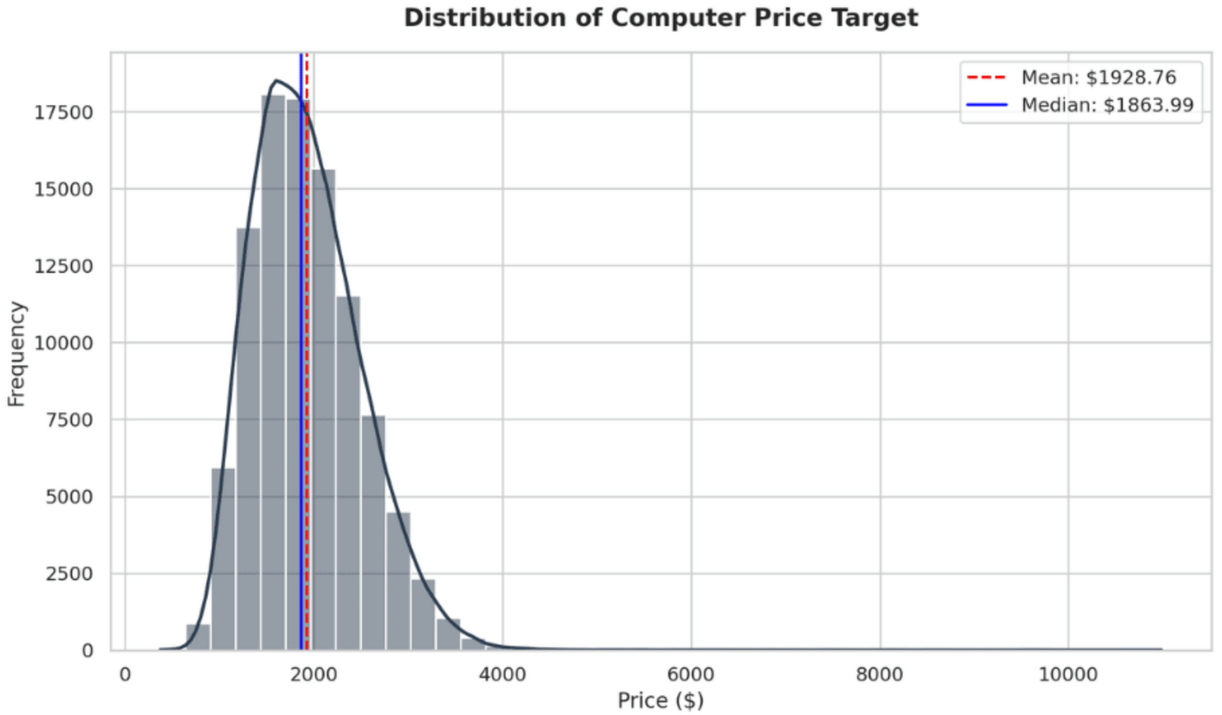
- Ekor panjang ke kanan
- Indikasi keberadaan premium devices

Mean price = $1,928.76

Median price = $1,863.99

Keberadaan ekor distribusi panjang menunjukkan adanya:

- Gaming laptop premium
- Workstation high-end
- Komputer custom dengan spesifikasi ekstrem



Kesimpulan

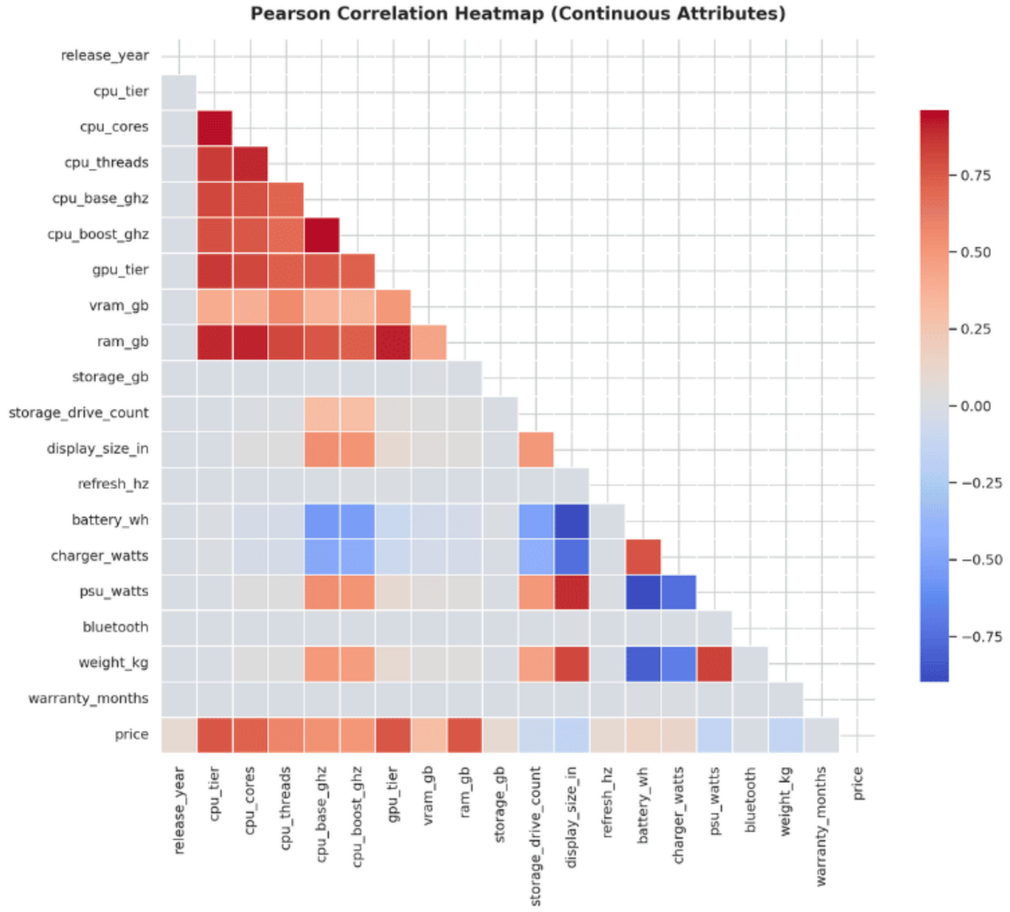
Distribusi harga memperlihatkan pasar komputer lebih banyak didominasi segmen mid-range dibanding ultra-premium.

FAKULTAS TEKNIK

Korelasi Spesifikasi Hardware

Presented by

Kelompok Keren



Hasil Korelasi

Heatmap Pearson menunjukkan RAM memiliki korelasi positif kuat terhadap harga, CPU dan GPU tier juga sangat berpengaruh, semakin tinggi spesifikasi, semakin tinggi harga

Korelasi Signifikan

ram_gb | cpu_tier | gpu_tier | cpu_cores | cpu_threads

Warna merah:
Korelasi positif kuat

Warna biru:
Korelasi negatif

Kesimpulan

Kapasitas RAM dan performa processor menjadi faktor dominan dalam pembentukan harga komputer modern.

FAKULTAS TEKNIK

Perbandingan Harga Brand

Presented by

Kelompok Keren

Hasil Boxplot

Visualisasi menunjukkan distribusi harga berbeda pada setiap brand komputer.
Brand premium seperti Apple, Razer memiliki median harga lebih tinggi dibanding brand lainnya.

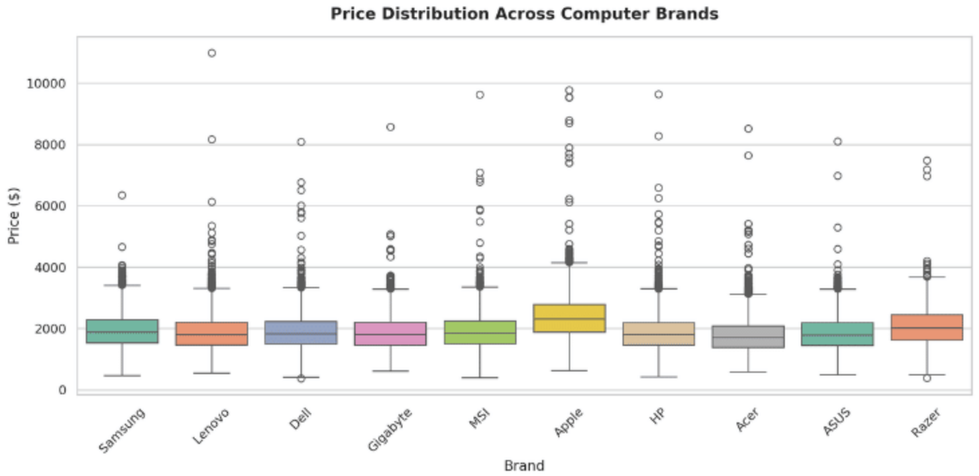
Segmentasi Pasar

Premium Segment
- Apple
- Razer

Karakteristik:
- Harga tinggi
- Target pasar eksklusif

Mainstream Segment
- HP
- Lenovo
- ASUS

Karakteristik:
- Pasar lebih luas
- Harga lebih kompetitif



Insight Analitik: Distribusi harga yang tidak normal menjadi indikasi preprocessing dan outlier handling sangat penting sebelum modelling.

Kesimpulan

Brand menjadi fitur penting dalam machine learning karena memiliki pengaruh signifikan terhadap harga jual komputer.

Data Preparation

Presented by

Kelompok Keren

FAKULTAS TEKNIK
UNIVERSITAS INDONESIA

Taksonomi Preprocessing Data

Presented by

Kelompok Keren

Pentingnya Preprocessing

Data mentah tidak dapat langsung digunakan untuk machine learning karena mengandung outlier, memiliki format kategorikal, distribusi data belum stabil

Tahapan Preprocessing

1. Raw Data
Dataset awal sebelum pembersihan.

2. Missing Value Audit
Memastikan tidak ada data kosong.

3. Outlier Handling
Menangani pencilan ekstrem.

4. Normalization & Encoding
Mengubah data menjadi format numerik standar.

5. Model Ready Dataset
Dataset siap digunakan untuk training model.

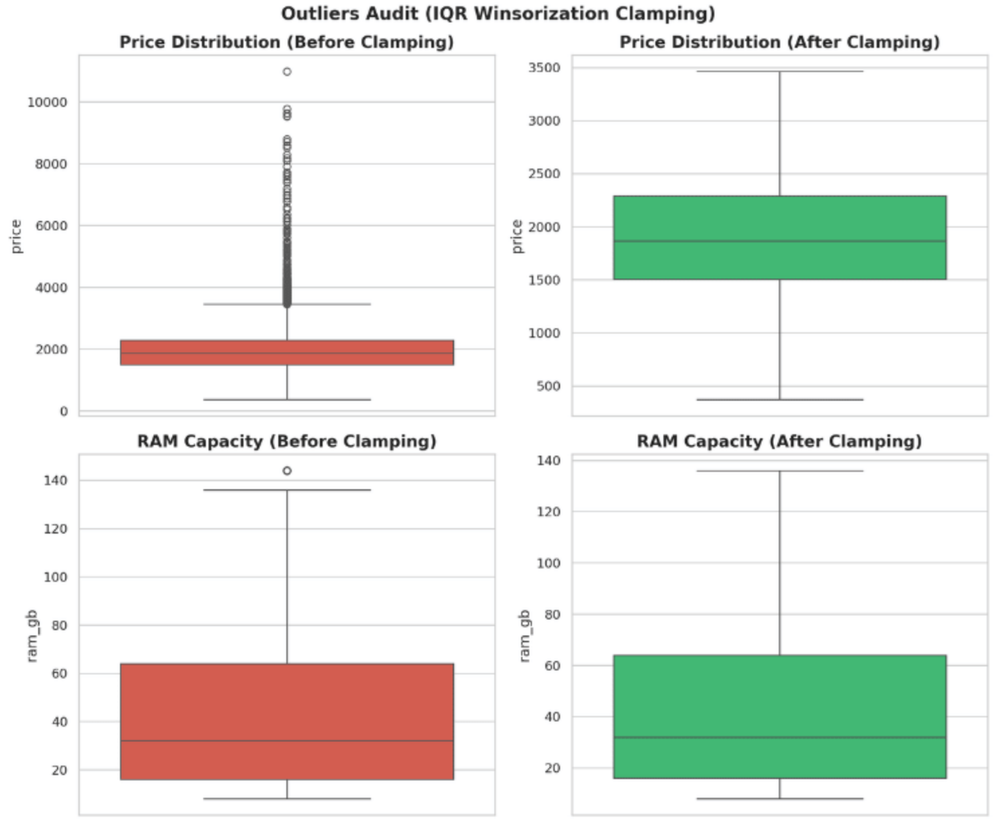
Tujuan Utama meningkatkan kualitas data, mengurangi noise, meningkatkan performa model

FAKULTAS TEKNIK

Identifikasi Outlier Dataset

Presented by

Kelompok Keren



Metode yang Digunakan

Deteksi outlier dilakukan menggunakan:

- Interquartile Range (IQR)

Metode ini mendeteksi data yang berada di luar:

- Q1 – 1.5(IQR)
- Q3 + 1.5(IQR)

Hasil Deteksi

Outlier terdeteksi pada:

Price = 976 data

RAM = 60 data

Kesimpulan

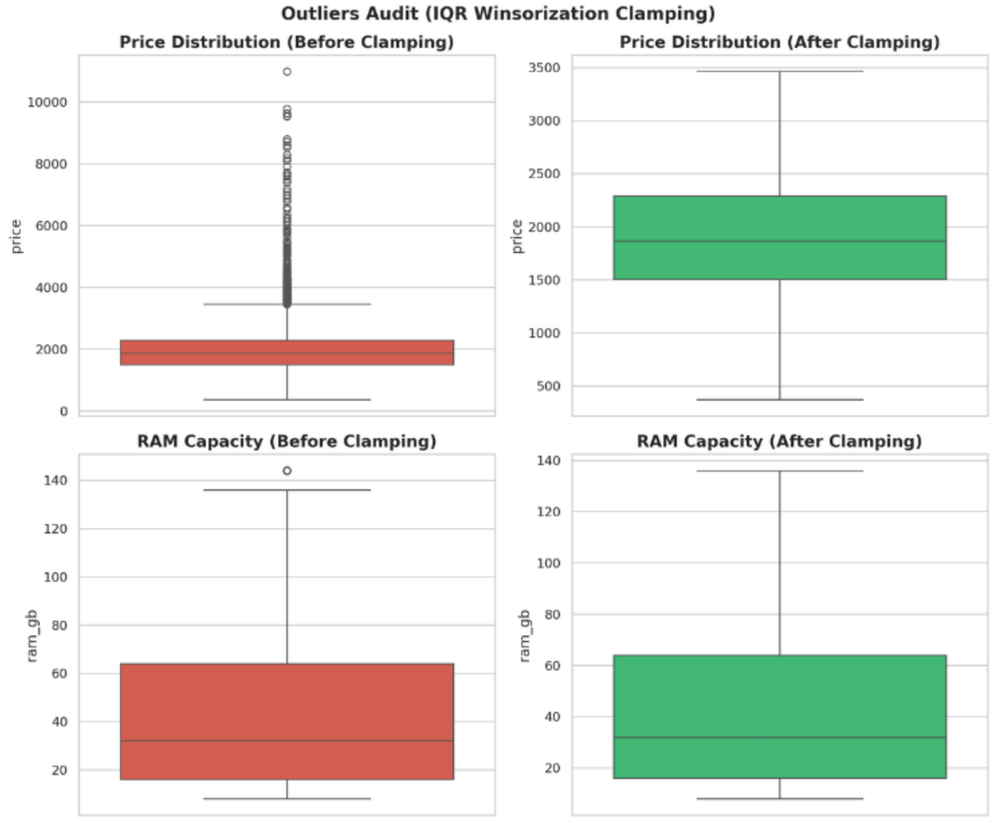
Outlier handling membantu meningkatkan stabilitas machine learning, akurasi prediksi, generalisasi model terhadap data baru

FAKULTAS TEKNIK

Penanganan Outlier Winsorization

Presented by

Kelompok Keren



Metode Penanganan Outlier

Setelah proses identifikasi outlier menggunakan metode IQR, dilakukan teknik:

- Winsorization
- Clamping data ekstrem
- Pembatasan nilai di luar ambang batas

Konsep Winsorization

Nilai ekstrem yang melebihi batas maksimum tidak dihapus, tetapi diturunkan ke batas ambang maksimum.

Hasil Setelah Clamping

Visualisasi setelah preprocessing menunjukkan outlier berhasil ditekan, distribusi menjadi lebih stabil

Kesimpulan

Pendekatan winsorization lebih efektif dibanding menghapus data karena tetap mempertahankan informasi penting tanpa merusak distribusi utama dataset.

FAKULTAS TEKNIK

Hubungan RAM dan Harga

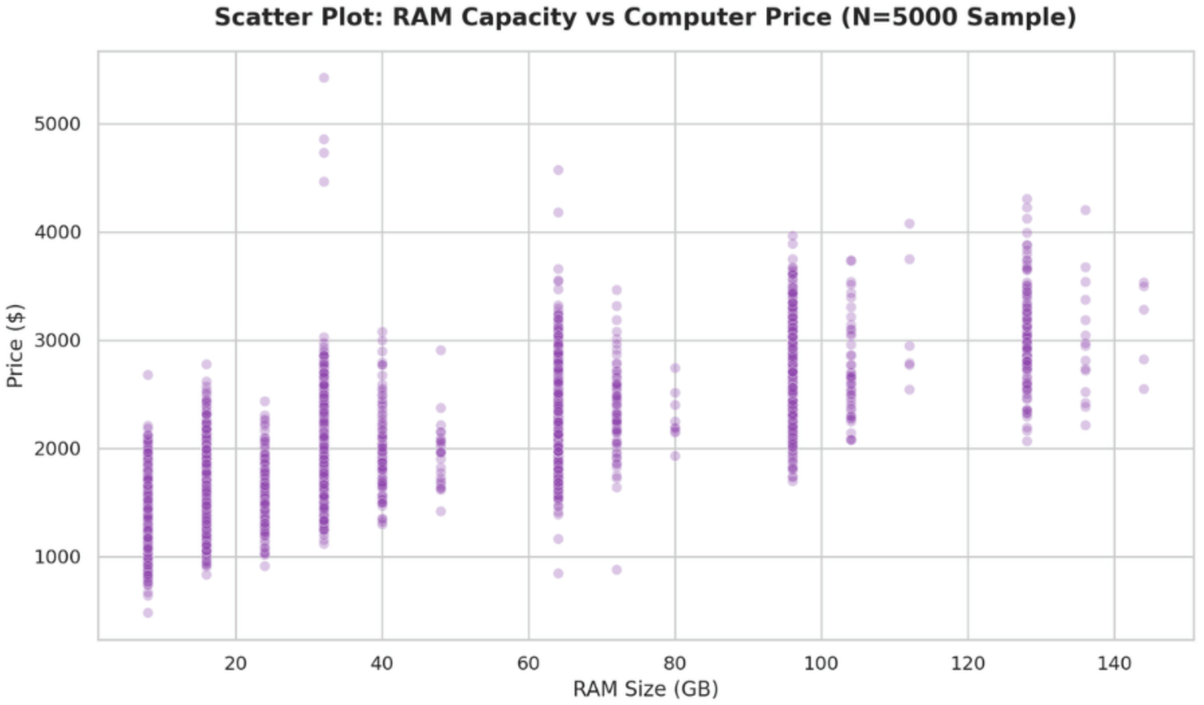
Presented by

Kelompok Keren

Hasil Visualisasi

Scatter plot menunjukkan:

- Tren linear positif yang kuat
- Harga meningkat seiring kapasitas RAM
- Sebaran data relatif konsisten



Interpretasi Analitik

Semakin besar kapasitas RAM:

- Semakin tinggi performa perangkat
- Semakin tinggi segmentasi pasar
- Semakin mahal harga jual komputer

Pola Distribusi

Perangkat dengan:

- RAM rendah → dominan pada harga budget
- RAM menengah → mendominasi pasar utama
- RAM tinggi → berada pada segmen premium

Kesimpulan

RAM menjadi salah satu fitur paling penting dalam menentukan estimasi harga komputer modern.

FAKULTAS TEKNIK

Densitas Berat dan Harga

Presented by

Kelompok Keren

Visualisasi Hexbin Density

Hexbin plot digunakan untuk memetakan kepadatan titik data, mengurangi overlap scatter plot, melihat konsentrasi distribusi harga

Hasil Analisis

Konsentrasi terbesar berada pada:

- Berat 1–3 kg
- Harga sekitar $1,500–$2,500

Area tersebut didominasi oleh:

- Laptop mainstream
- Perangkat portable

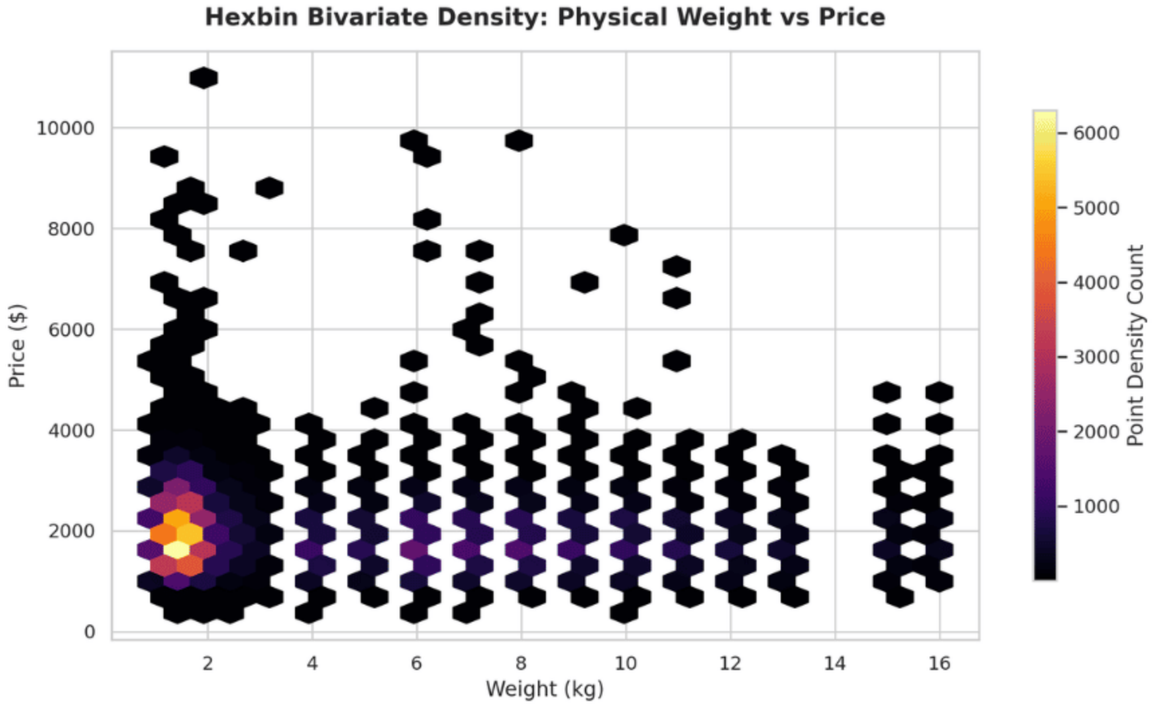
Segmentasi Berat

Lightweight Devices:

- Laptop tipis
- Mobilitas tinggi
- Harga menengah

Heavyweight Devices

- Desktop tower
- Gaming workstation
- Harga premium



Kesimpulan

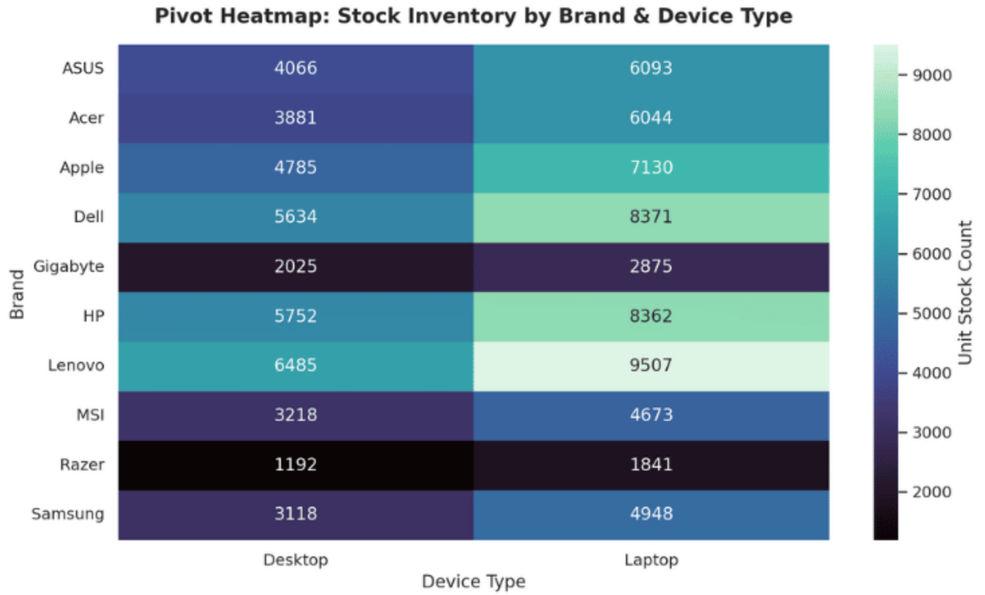
Hexbin density membantu memperlihatkan pola distribusi pasar komputer berdasarkan kombinasi berat fisik dan harga jual.

FAKULTAS TEKNIK

Pivot Harga Brand

Presented by

Kelompok Keren



Kesimpulan

Pivot heatmap membantu memahami hubungan antara Brand, Device type, Average selling price secara lebih visual dan terstruktur.

Tujuan Pivot Table

Pivot table digunakan untuk membandingkan rata-rata harga, menganalisis segmentasi pasar, mengidentifikasi brand premium

Hasil Analisis

Brand Apple memiliki:

- Rata-rata harga tertinggi
- Dominasi pada segmen premium
- Konsistensi harga tinggi pada desktop maupun laptop

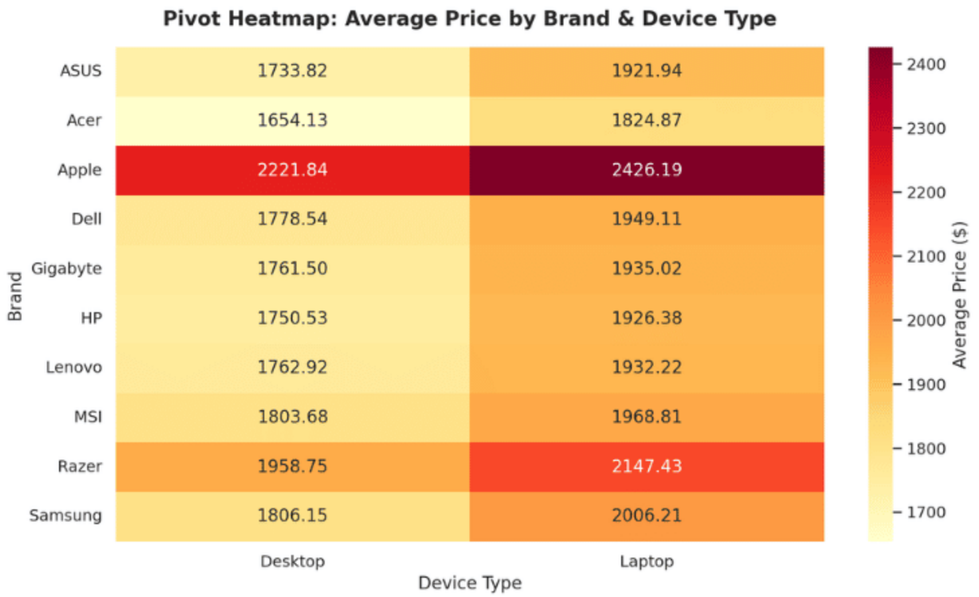
Premium	Mainstream	Budget
Apple, Razer	Dell, HP, Lenovo, ASUS	Acer

FAKULTAS TEKNIK

Distribusi Volume Inventori

Presented by

Kelompok Keren



Kesimpulan

Distribusi volume inventori membantu memahami struktur pasar komputer berdasarkan brand dan form factor perangkat.

Tujuan Analisis

Pivot distribution digunakan untuk menghitung jumlah unit tiap brand, membandingkan volume desktop dan laptop, mengukur dominasi pasar

Hasil Visualisasi

Volume terbesar ditemukan pada **Lenovo Laptop, Dell Laptop, HP Laptop.** Hal ini menunjukkan dominasi laptop pasar modern.

Interpretasi Pasar

Laptop memiliki permintaan lebih tinggi, distribusi pasar lebih luas, segmentasi pengguna lebih besar

FAKULTAS TEKNIK

Pengantar Feature Engineering

Presented by

Kelompok Keren

Definisi

Feature engineering merupakan proses transformasi data, rekayasa atribut, dan pembuatan fitur baru untuk meningkatkan performa model machine learning.

Tujuan Utama

- Mempermudah model memahami pola
- Menyederhanakan data kompleks
- Meningkatkan kualitas prediksi

Proses Transformasi

Sebelum Engineering

- Data mentah
- Variabel kompleks
- Format belum optimal

Setelah Engineering

- Data lebih informatif
- Fitur lebih relevan
- Struktur lebih mudah dipelajari model

Kesimpulan

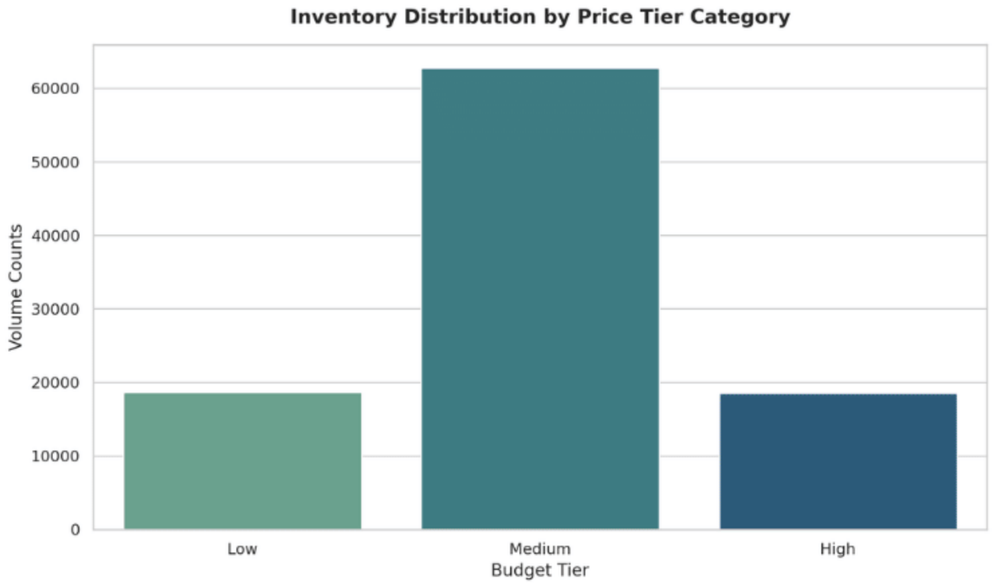
Feature engineering menjadi salah satu tahap paling penting dalam membangun model machine learning berkualitas tinggi.

FAKULTAS TEKNIK

Kategorisasi Price Binning

Presented by

Kelompok Keren



Kesimpulan

Feature binning membantu mengubah data kontinu menjadi representasi kategori yang lebih mudah dianalisis.

Konsep Price Binning

Variabel harga kontinu disederhanakan menjadi Low, Medium, dan High berdasarkan interval tertentu.

Tujuan Binning

- Menyederhanakan distribusi harga
- Mempermudah segmentasi pasar
- Membantu proses filtering data

Segmentasi Kategori

Low Tier: Harga budget, entry-level devices

Medium Tier: mainstream market, mid-range devices

High Tier: premium computers, gaming workstation

FAKULTAS TEKNIK

One-Hot Encoding Nominal

Presented by

Kelompok Keren

Permasalahan Data Kategorikal

Model machine learning tidak dapat langsung membaca data teks, variabel string, dan kategori nominal

Contoh Transformasi

Kolom:
- device_type

Diubah menjadi:
- device_desktop
- device_laptop

Solusi One-Hot Encoding

Variabel kategorikal diubah menjadi:
- Dummy variables
- Representasi biner
- Nilai 0 dan 1

Keunggulan Teknik

- Tidak menciptakan ranking palsu
- Aman untuk data nominal
- Mudah diproses model regresi

Kesimpulan

One-Hot Encoding memungkinkan data kategorikal diproses menjadi format numerik yang kompatibel dengan machine learning.

FAKULTAS TEKNIK

Ordinal Encoding Tingkatan

Presented by

Kelompok Keren

Definisi dan Tujuan Encoding

Ordinal Encoding digunakan untuk data kategorikal bertingkat dan variabel dengan urutan logis serta kategori berhierarki. Tujuan Encoding untuk mempertahankan hubungan ranking, mempermudah pemrosesan numerik, menyederhanakan fitur kategorikal.

Contoh Encoding

Kategori:

- Low → 0
- Medium → 1
- High → 2

One-Hot Encoding

Digunakan untuk:

- Data nominal
- Tanpa urutan kategori

Ordinal Encoding

Digunakan untuk:

- Data bertingkat
- Memiliki relasi ranking

Kesimpulan

Ordinal encoding membantu model memahami tingkatan kategori harga secara matematis dan terstruktur.

FAKULTAS TEKNIK

Konsep Seleksi Fitur

Presented by

Kelompok Keren

Definisi

Feature selection merupakan proses memilih fitur paling relevan untuk digunakan pada model machine learning.

Tujuan Utama & Dampak

- Mengurangi dimensi data
- Menghapus fitur redundan
- Mempercepat komputasi model
- Mengurangi overfitting

Feature selection membantu:

- Meningkatkan efisiensi model
- Menjaga stabilitas prediksi
- Mengurangi multicollinearity

Filter Method

Berbasis:

- Statistik
- Korelasi
- Hubungan antar variabel

Wrapper Method

Berbasis:

- Iterasi model
- Recursive feature elimination

Embedded Method

Berbasis:

- Seleksi otomatis dari algoritma model

Kesimpulan

Seleksi fitur merupakan tahap penting untuk menghasilkan model machine learning yang lebih ringan, cepat, dan akurat.

Feature Selection

Presented by

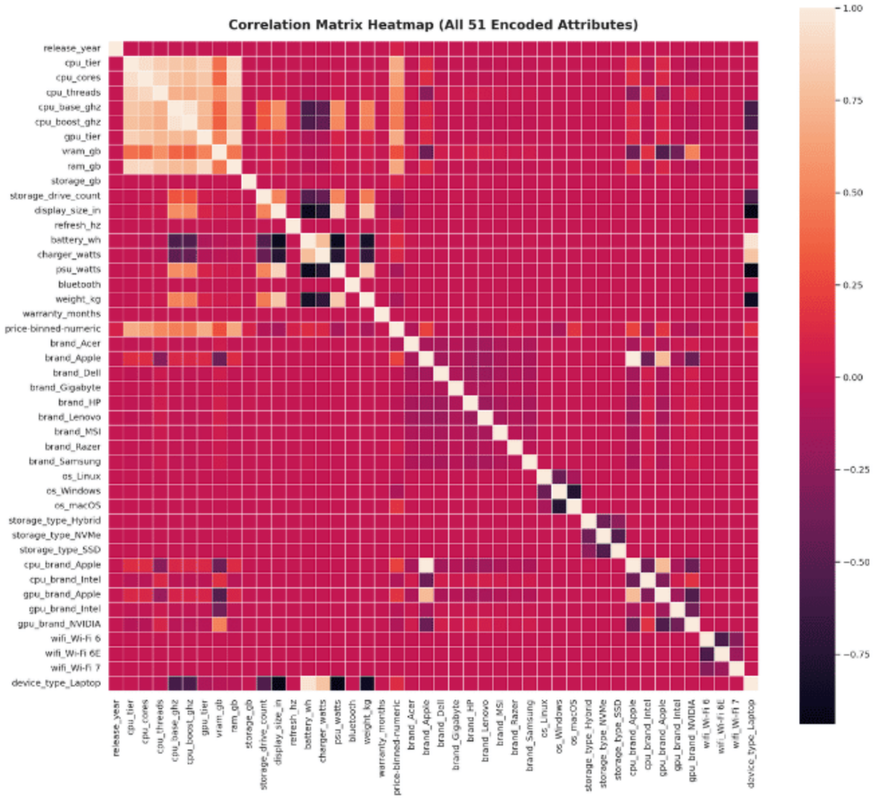
Kelompok Keren

FAKULTAS TEKNIK
UNIVERSITAS INDONESIA

Korelasi Fitur Encoding

Presented by

Kelompok Keren



Kesimpulan

Correlation heatmap membantu mengidentifikasi fitur redundan sebelum proses feature selection dilakukan.

Matriks Korelasi Full Encoding

Heatmap menampilkan 51 variabel hasil encoding, hubungan antar fitur numerik dan kategorikal, dan tingkat korelasi antar atribut hardware

Interpretasi Warna Heatmap

- Warna terang → korelasi tinggi
- Warna gelap → korelasi rendah
- Korelasi mendekati 1 → potensi redundansi data

Dampak terhadap Model

Korelasi berlebih dapat menyebabkan overfitting, ketidakstabilan koefisien model, dan penurunan generalisasi prediksi

FAKULTAS TEKNIK

Filter Method Korelasi

Presented by

Kelompok Keren

Konsep Filter Method

Filter Method bekerja menggunakan statistik korelasi, hubungan terhadap target, dan eliminasi fitur redundan.

Strategi Seleksi

Tahapan seleksi:

- Menghitung korelasi seluruh fitur
- Menentukan ambang batas minimum
- Menghapus fitur dengan redundansi tinggi

Hasil Seleksi

Kriteria utama:

- Korelasi absolut terhadap price ≥ 0.1
- Tidak memiliki multikolinearitas tinggi

Dampak Seleksi

Fitur yang tidak relevan:

- Dihapus dari dataset
- Mengurangi dimensi data
- Mempercepat training model

Kesimpulan

Filter Method menghasilkan dataset yang lebih ringan dengan tetap mempertahankan fitur yang memiliki kontribusi signifikan terhadap prediksi harga.

FAKULTAS TEKNIK

Wrapper Method RFE

Presented by

Kelompok Keren

Konsep Wrapper Method

Wrapper Method bekerja secara iteratif dengan melatih model berulang kali, mengevaluasi kontribusi fitur, dan menghapus fitur terlemah bertahap

Algoritma yang Digunakan

Metode menggunakan:

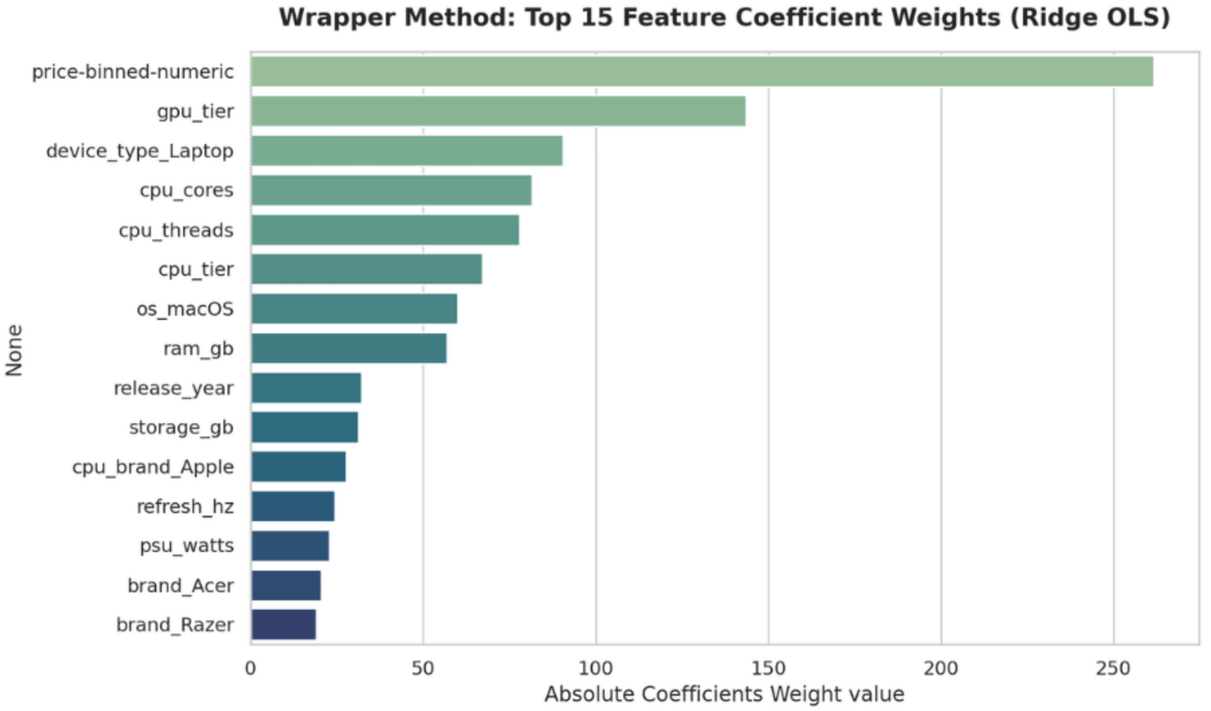
- Ridge Regression
- Recursive Feature Elimination (RFE)

untuk memilih fitur paling optimal.

Hasil Seleksi

Terpilih 15 fitur terbaik dengan bobot kontribusi tertinggi, seperti:

- gpu_tier
- ram_gb
- device_type_Laptop
- cpu_threads
- os_macOS



Kesimpulan

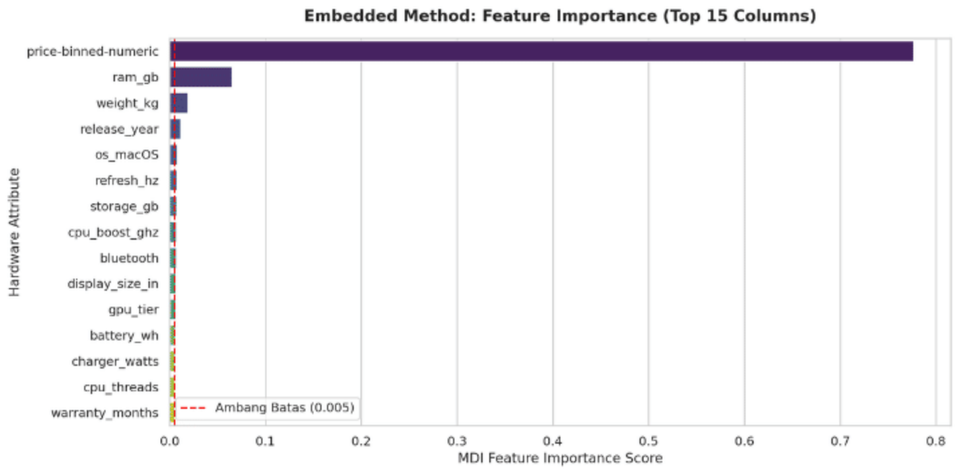
Wrapper Method menghasilkan kombinasi fitur paling optimal dengan performa prediksi terbaik dibanding metode lainnya.

FAKULTAS TEKNIK

Embedded Feature Importance

Presented by

Kelompok Keren



Kesimpulan

Embedded Method mampu memilih fitur penting secara otomatis menggunakan mekanisme internal model Random Forest.

Konsep Embedded Method

Embedded Method melakukan seleksi fitur otomatis, penilaian langsung dari model, dan evaluasi kontribusi intrinsik fitur.

Algoritma Digunakan

Model: Random Forest Regressor
Mengukur Mean Decrease in Impurity (MDI) dan tingkat kontribusi fitur pada pohon keputusan

Hasil Analisis

Fitur paling dominan yaitu price-binned-numeric, ram_gb, weight_kg, dan release_year

FAKULTAS TEKNIK

Evaluasi Seleksi Fitur

Presented by

Kelompok Keren

Tujuan Evaluasi

Seluruh metode feature selection diuji menggunakan KNN Regressor, dataset test set identik, dan parameter evaluasi yang sama

Metrik Evaluasi

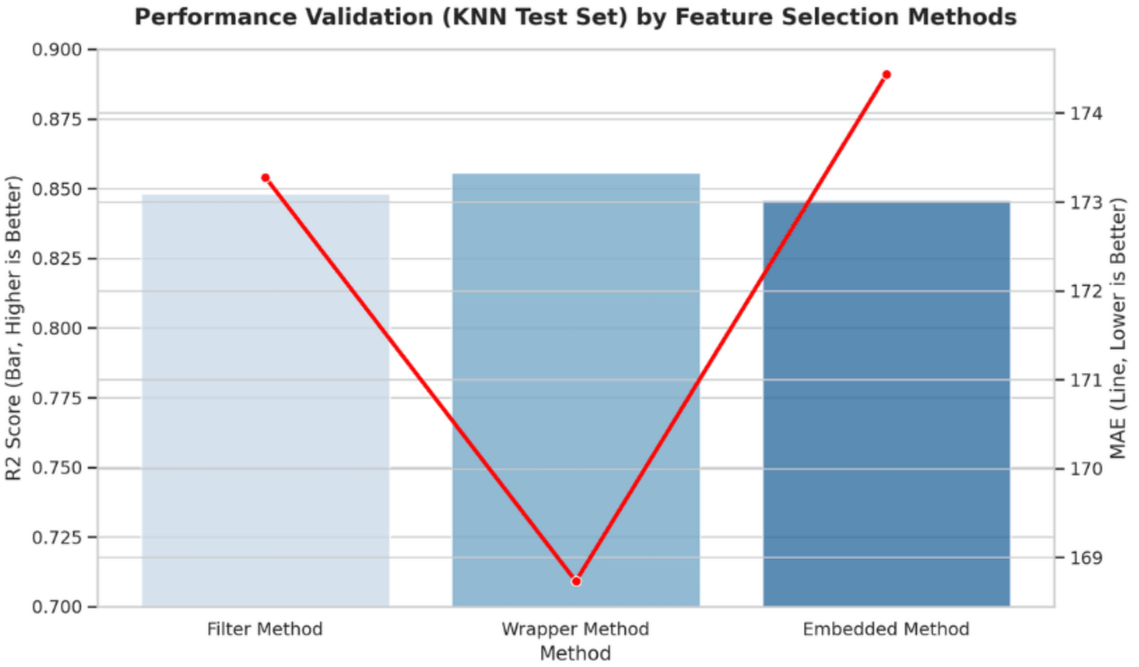
Digunakan dua indikator utama:

- R^2 Score → semakin tinggi semakin baik
- MAE → semakin rendah semakin baik

Hasil Terbaik

Wrapper Method menjadi pemenang karena:

- Memiliki R^2 tertinggi
- Memiliki MAE terkecil
- Memberikan generalisasi model terbaik



Kesimpulan

Wrapper Method dipilih sebagai dataset final karena menghasilkan performa prediksi paling optimal.

FAKULTAS TEKNIK
UNIVERSITAS INDONESIA

Dataset Final Pemodelan

Presented by

Kelompok Keren

Dataset Terpilih	Perubahan Dimensi Data	Dampak Optimasi
Dataset final yang digunakan: • df_wrapper.csv berasal dari hasil seleksi Wrapper Method.	**Dataset Awal** • 100.000 baris • 51 fitur **Dataset Final** • 100.000 baris • 15 fitur optimal	**Reduksi fitur menghasilkan:** • Komputasi lebih cepat • Training lebih ringan • Efisiensi memori lebih baik

Keunggulan Dataset Final

Dataset hasil wrapper:

Lebih bersih	Minim redundansi	Fokus pada fitur paling informatif

Kesimpulan

Feature selection berhasil menyederhanakan dataset secara signifikan tanpa mengorbankan akurasi model.

Modelling & Evaluasi

Presented by

Kelompok Keren

FAKULTAS TEKNIK

Konsep Supervised Learning

Presented by

Kelompok Keren

Strategi Pemodelan

Dilakukan pengujian terhadap:
- 5 algoritma regresi berbeda
- Evaluasi objektif performa
- Pemilihan model terbaik

Arsitektur Supervised Machine Learning Regresi

Supervised learning merupakan metode pembelajaran berbasis label, memetakan input ke output,dan belajar dari data historis

Classification

Output:
- Kategori
- Label diskret

Contoh:
- Spam / Non-spam

Regression

Output:
- Nilai kontinu
- Angka numerik

Contoh:
- Prediksi harga komputer

Pendekatan Penelitian

Karena target berupa harga numerik, maka:
- Tipe masalah = Regression
- Fokus utama = estimasi nilai price

Kesimpulan

Supervised regression dipilih karena paling sesuai untuk memprediksi nilai harga komputer berbasis spesifikasi hardware.

FAKULTAS TEKNIK

Model KNN Regressor

Presented by

Kelompok Keren

Konsep Dasar KNN dan Perhitungan Jarak

KNN bekerja dengan mencari tetangga terdekat, mengukur kemiripan spesifikasi, mengambil rata-rata harga sekitar. Model menggunakan Euclidean Distance untuk menghitung kedekatan antar komputer.

Cara Kerja

Setiap fitur memiliki:

- Bobot kontribusi
- Pengaruh positif atau negatif
- Nilai koefisien berbeda

Keunggulan Model

- Mudah diinterpretasikan
- Cepat dilatih
- Cocok untuk hubungan linear

Kelemahan Model

- Sulit menangkap pola non-linear
- Sensitif terhadap multikolinearitas
- Asumsi statistik cukup ketat

Kesimpulan

Regresi linier menjadi model fundamental untuk memahami hubungan matematis antar spesifikasi komputer dan harga pasar.

FAKULTAS TEKNIK

Regresi Linier Berganda

Presented by

Kelompok Keren

Definisi Model dan Konsep Matematis

Regresi linier bergandamembentuk persamaan matematis global dan menghubungkan banyak fitur dengan target harga. Model menghitung koefisien β tiap fitur, hubungan linear terhadap price, serta pengaruh masing-masing atribut hardware

Cara Kerja Model

Tahapan:

1. Data baru dimasukkan
2. Sistem mencari k tetangga terdekat
3. Harga diprediksi dari rata-rata tetangga

Keunggulan KNN

- Sederhana
- Tidak memerlukan training kompleks
- Efektif pada pola lokal data

Kelemahan KNN

- Sensitif terhadap outlier
- Berat pada dataset besar
- Lambat saat inferensi

Kesimpulan

KNN cocok digunakan sebagai baseline model dalam membandingkan performa algoritma regresi lainnya.

FAKULTAS TEKNIK

Support Vector Regressor

Presented by

Kelompok Keren

Konsep Dasar SVR dan Margin Epsilon

SVR bekerja dengan membentuk garis regresi optimal dan meminimalkan error prediksi serta mengontrol margin toleransi. Model menggunakan ϵ-tube (epsilon tube) untuk menentukan batas toleransi error yang masih dianggap aman.

Mekanisme Model

SVR fokus pada:

- Titik data di luar margin
- Support vectors
- Optimasi hyperplane regresi

Keunggulan SVR

- Stabil terhadap noise
- Generalisasi tinggi
- Efektif pada data kompleks

Kelemahan SVR

- Training lebih lambat
- Sensitif terhadap tuning parameter
- Kompleks pada dataset besar

Kesimpulan

KNN cocok digunakan sebagai baseline model dalam membandingkan performa algoritma regresi lainnya.

FAKULTAS TEKNIK
UNIVERSITAS INDONESIA

Decision Tree & Random Forest

Presented by

Kelompok Keren

Decision Tree Regressor

Decision Tree bekerja dengan:
- Membagi data secara bertahap
- Menggunakan logika IF–THEN
- Membentuk struktur pohon keputusan

Karakteristik Decision Tree

Keunggulan:
- Mudah dipahami
- Interpretatif
- Mampu menangkap pola non-linear

Kelemahan:
- Mudah overfitting
- Sensitif terhadap noise data

Random Forest Regressor

Random Forest merupakan:
- Ensemble Learning
- Gabungan banyak Decision Tree
- Menggunakan teknik Bagging

Cara Kerja Random Forest

Tahapan:
- Membuat banyak pohon keputusan
- Setiap pohon dilatih secara independen
- Hasil prediksi dirata-ratakan

FAKULTAS TEKNIK

Metrik Evaluasi Regresi

Presented by

Kelompok Keren

Tujuan Evaluasi

Evaluasi dilakukan untuk mengukur akurasi prediksi, mengukur besar kesalahan model, membandingkan performa antar algoritma

R-Squared (R^2)

R^2 mengukur:
- Persentase variansi target
- Kemampuan model menjelaskan data

Interpretasi:
- Semakin mendekati 1 → semakin baik

Mean Absolute Error (MAE)

MAE mengukur:
- Rata-rata selisih absolut
- Error prediksi terhadap nilai aktual

Interpretasi:
- Semakin kecil → semakin akurat

Root Mean Squared Error (RMSE)

RMSE:
- Memberi penalti besar pada error ekstrem
- Sensitif terhadap outlier

FAKULTAS TEKNIK

GridSearchCV & Cross Validation

Presented by

Kelompok Keren

Tujuan Optimasi

Optimasi dilakukan untuk mencari parameter terbaik, menghindari parameter manual, meningkatkan generalisasi model

GridSearchCV

GridSearchCV bekerja dengan:

- Menguji seluruh kombinasi parameter
- Membandingkan performa otomatis
- Memilih konfigurasi terbaik

Cross Validation

Menggunakan:

- 5-Fold Cross Validation

Data dibagi menjadi:

- 5 subset berbeda
- Training dan validation bergantian

Keunggulan Pendekatan

Cross validation:

- Mengurangi bias evaluasi
- Menguji kestabilan model
- Memastikan performa konsisten

FAKULTAS TEKNIK
UNIVERSITAS INDONESIA

Perbandingan Akurasi Model

Presented by

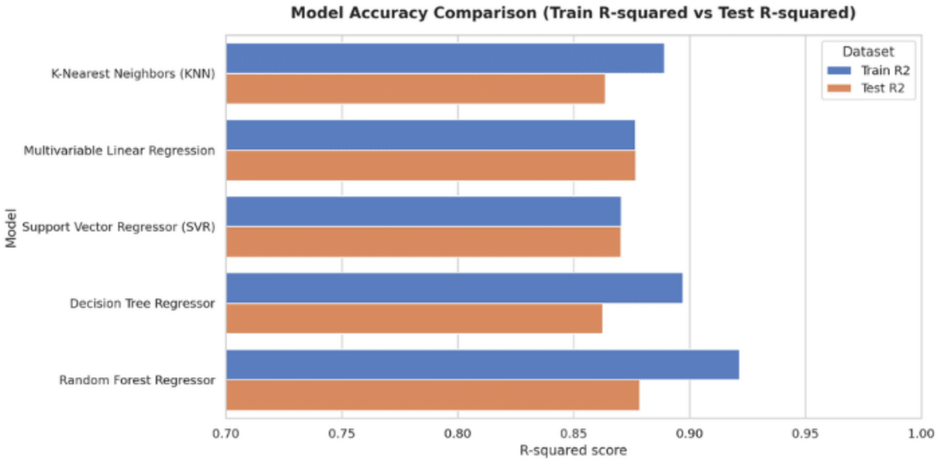
Kelompok Keren

Tujuan Analisis

Membandingkan akurasi training, akurasi testing, kemampuan generalisasi model

Hasil Performa Model

Model	Test R^2
KNN	~0.864
Linear Regression	~0.877
SVR	~0.870
Decision Tree	~0.863
Random Forest	~0.879



Insight Utama: Random Forest memiliki R^2 testing tertinggi, menunjukkan akurasi paling baik, konsisten pada data baru

Kesimpulan

Random Forest menjadi model dengan performa prediksi paling unggul dibanding algoritma lainnya.

FAKULTAS TEKNIK
UNIVERSITAS INDONESIA

Perbandingan Error Model

Presented by

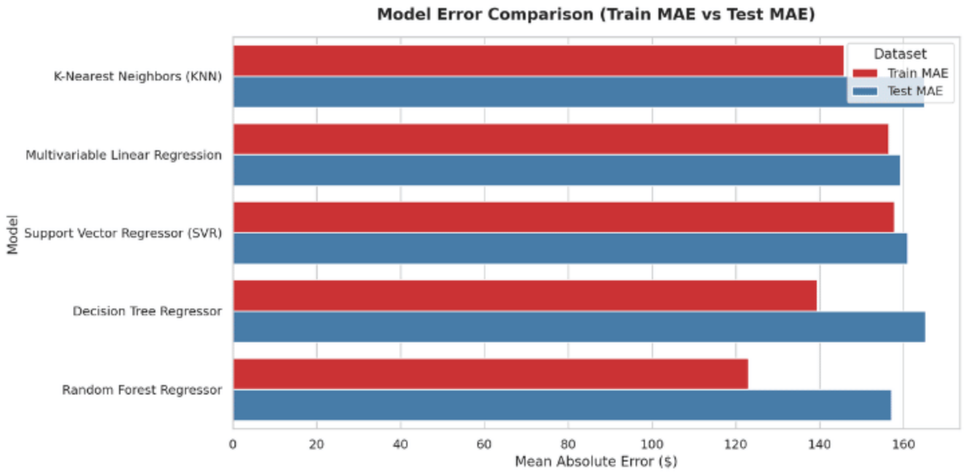
Kelompok Keren

Tujuan Analisis

MAE digunakan untuk mengukur rata-rata eror prediksi dan menilai ketepatan model regresi.

MAE Testing

Model	Test R^2
KNN	~165
Linear Regression	~159
SVR	~161
Decision Tree	~165
Random Forest	~157



Insight Utama: Random Forest memiliki error terkecil, prediksi paling presisi, stabil pada berbagai rentang harga

Kesimpulan

Random Forest menghasilkan keseimbangan terbaik antara akurasi tinggi dan error rendah.

FAKULTAS TEKNIK

Pemilihan Model Terbaik

Presented by

Kelompok Keren

Hasil Kompetisi Model

Berdasarkan seluruh pengujian, Random Forest menjadi model terbaik karena mampu mengungguli model lain data testing.

Performa Akhir		Alasan dan Kelebihan	
Test R^2	87.85%	Random Forest unggul karena: • Akurasi tinggi • Error rendah • Stabil terhadap data baru	**Model:** • Cocok untuk implementasi nyata • Stabil pada skala besar • Adaptif terhadap variasi spesifikasi hardware
Test MAE	$157.13		

Kesimpulan

Random Forest dipilih sebagai model final untuk sistem prediksi harga komputer otomatis.

FAKULTAS TEKNIK

Scatter Plot
Aktual vs Prediksi

Presented by

Kelompok Keren

Tujuan Scatter Plot

Scatter plot digunakan untuk membandingkan nilai aktual vs prediksi, mengukur kedekatan hasil model, memvisualisasikan akurasi

Interpretasi Grafik

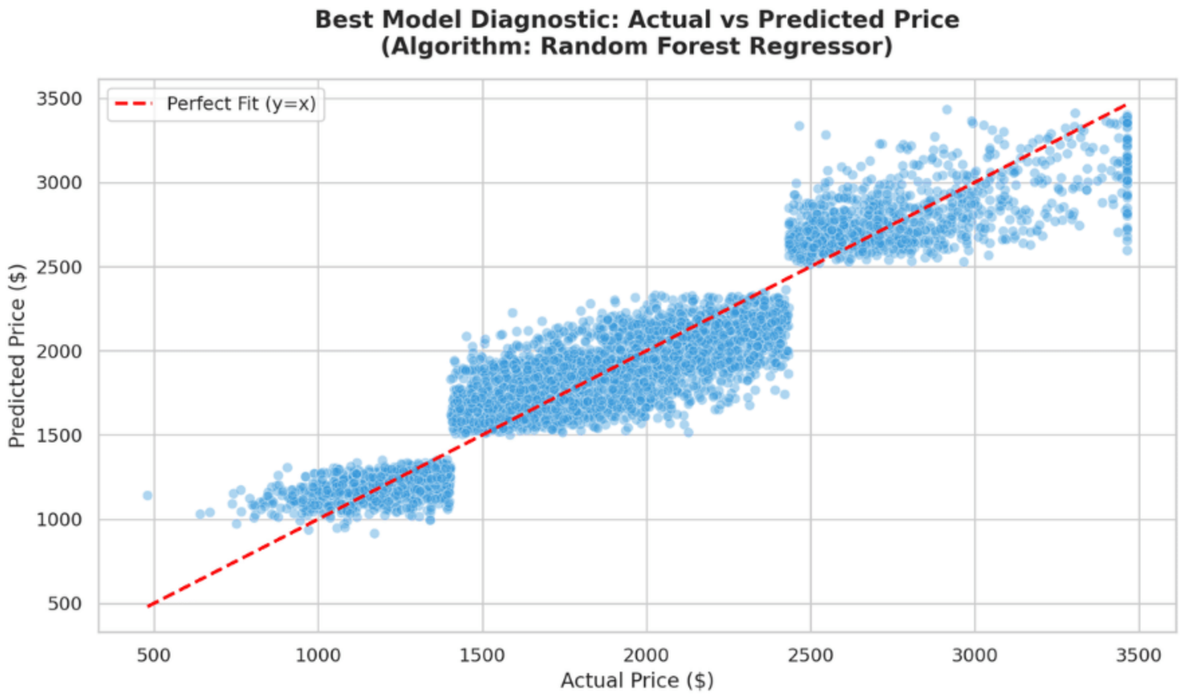
Garis merah diagonal:

- Menunjukkan prediksi sempurna
- Semakin dekat titik ke garis → semakin akurat

Temuan Utama

Sebagian besar titik:

- Mengelompok dekat garis identitas
- Menandakan error kecil
- Menunjukkan prediksi konsisten



Kesimpulan

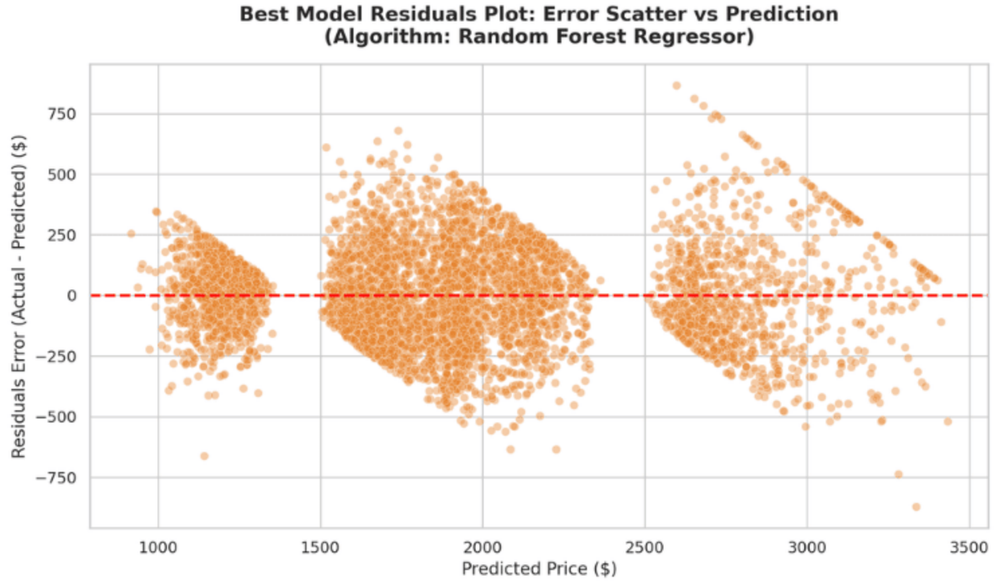
Scatter plot membuktikan bahwa model Random Forest mampu menghasilkan prediksi yang sangat mendekati harga aktual pasar.

FAKULTAS TEKNIK

Residual Plot

Presented by

Kelompok Keren



Tujuan Residual Plot

Residual plot digunakan untuk mengevaluasi pola error, menguji kestabilan model, mendeteksi heteroskedastisitas

Interpretasi Residual

Residual:

- Actual – Predicted
- Error positif → model underpredict
- Error negatif → model overpredict

Temuan Utama

Sebaran residual acak di sekitar nol, tidak membentuk pola tertentu

Analisis Statistik

Tidak terlihat pola corong ekstrem, penyimpangan besar berulang

Kesimpulan

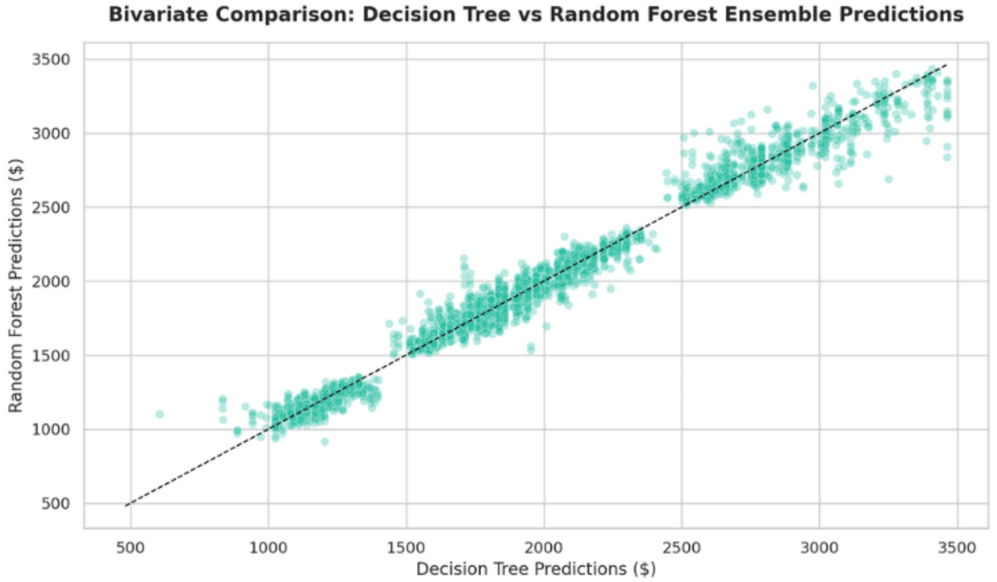
Correlation heatmap membantu mengidentifikasi fitur redundan sebelum proses feature selection dilakukan.

FAKULTAS TEKNIK
UNIVERSITAS INDONESIA

Decision Tree vs Random Forest

Presented by

Kelompok Keren



Tujuan Perbandingan

Membandingkan:
- Prediksi Decision Tree tunggal
- Prediksi Random Forest ensemble

Efek Ensemble

Random Forest:
- Merata-ratakan banyak pohon
- Mengurangi noise
- Menekan overfitting

Decision Tree

- Prediksi lebih kasar
- Variansi lebih tinggi

Random Forest

- Prediksi lebih halus
- Distribusi lebih stabil

Kesimpulan

Random Forest lebih unggul dibanding Decision Tree tunggal karena mampu menghasilkan prediksi yang lebih konsisten dan robust.

FAKULTAS TEKNIK

Diagnosis Overfitting & Penutup

Presented by

Kelompok Keren

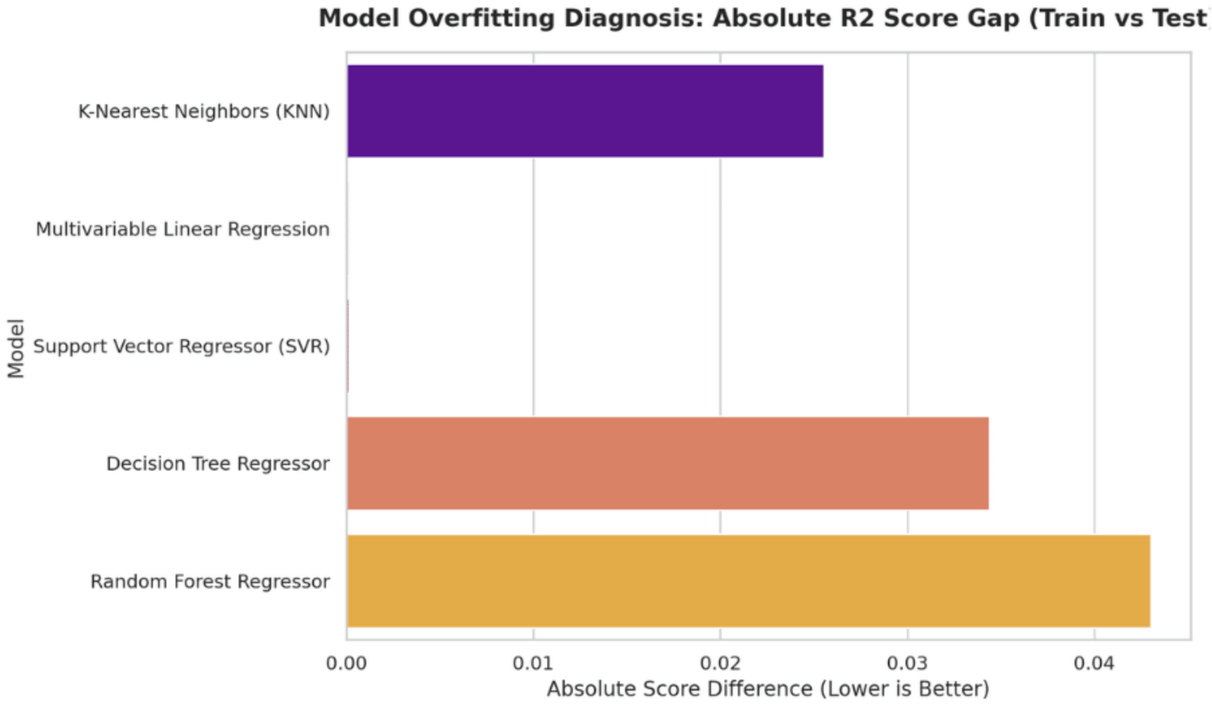
Analisis Overfitting

Overfitting dianalisis menggunakan:

- Selisih Train R^2 vs Test R^2
- Konsistensi performa model

Hasil Diagnosis

Model	Gap R^2
KNN	Kecil
Linear Regression	Sangat kecil
SVR	Hampir nol besar
Decision Tree	Lebih besar
Random Forest	Tetap stabil



Insight Utama

Random Forest tidak mengalami overfitting berat, tetap stabil pada data testing, memiliki generalisasi sangat baik

FAKULTAS TEKNIK

Kesimpulan Akhir Penelitian

Presented by

Kelompok Keren

Hasil Utama

Penelitian berhasil:

- Membangun sistem prediksi harga komputer
- Menggunakan pendekatan machine learning regresi
- Mencapai akurasi tinggi

Model Terbaik

Random Forest Regressor:

- R^2 = 87.85%
- MAE = $157.13

menjadi model paling optimal.

Kontribusi Penelitian

Sistem:

- Membantu estimasi harga otomatis
- Mendukung pengambilan keputusan bisnis
- Mengurangi subjektivitas penilaian harga

Terima Kasih!

Presented by

Kelompok Keren