



Analisis Faktor Penentu Profit Penjualan Mobil Menggunakan Algoritma Random Forest

Muhamad Fahrul Rozi*, Mukhammad Fakhir Rizal

Program Studi Sistem Informasi Industri Otomotif, Politeknik STMI Jakarta, Jakarta Pusat, Indonesia

Email: ^{1,*}fahrul@stmi.ac.id, ²fakhir_rizal@stmi.ac.id

Email Penulis Korespondensi: fahrul@stmi.ac.id

Abstrak—Industri otomotif mengalami perubahan signifikan dalam beberapa tahun terakhir yang berdampak langsung pada profitabilitas penjualan mobil. Tujuan dari studi ini adalah untuk menganalisis faktor-faktor yang memengaruhi profit penjualan mobil menggunakan *dataset USA car sales* periode 2018–2024. Pendekatan yang digunakan adalah metode kuantitatif berbasis *machine learning* dengan algoritma *random forest*, yang dipilih karena kemampuannya dalam menangani data kompleks serta mengidentifikasi variabel-variabel penting yang berkontribusi terhadap profit. Analisis dilakukan melalui beberapa tahapan, meliputi pemrosesan data, pelatihan model, evaluasi kinerja, serta interpretasi hasil menggunakan teknik *feature importance*. Tahapan tersebut bertujuan untuk memperoleh model yang akurat sekaligus memberikan pemahaman yang komprehensif mengenai pengaruh setiap variabel terhadap profit penjualan mobil. Hasil analisis menunjukkan bahwa beberapa faktor memiliki pengaruh signifikan terhadap profit penjualan mobil, di antaranya merek mobil, tahun dilakukannya transaksi penjualan, dan jumlah unit mobil yang dibeli dalam satu transaksi. Merek mobil mencerminkan preferensi pasar dan segmentasi konsumen, sementara tahun penjualan menggambarkan dinamika tren dan kondisi pasar. Selain itu, jumlah unit yang terjual dalam satu transaksi turut berperan penting dalam meningkatkan total profit. Temuan ini memberikan wawasan strategis bagi perusahaan otomotif dalam merumuskan kebijakan dan strategi penjualan yang lebih efektif, adaptif terhadap perubahan pasar, serta berbasis pada analisis data.

Kata Kunci: Intelijen Bisnis; Analisis Kuantitatif; Algoritma Random Forest; Analisis Penjualan Mobil; Kinerja Penjualan

Abstract—The automotive industry has significant changes in recent years that have directly affected vehicle sales profitability. The objective of this study is to analyze the factors influencing car sales profit using the USA Car Sales dataset for the 2018–2024 period. The approach employed is a quantitative method based on machine learning using the random forest algorithm, which was selected for its ability to handle complex data and identify important variables contributing to profit. The analysis was conducted through several stages, including data preprocessing, model training, performance evaluation, and result interpretation using feature importance techniques. These stages aim to obtain an accurate model while providing a comprehensive understanding of the influence of each variable on car sales profit. The results indicate that several factors have a significant impact on car sales profit, including car brand, year of sale, and the number of units purchased in a single transaction. Car brand reflects market preferences and consumer segmentation, while the year of sale represents market trends and changing conditions over time. In addition, the number of units sold per transaction plays an important role in increasing total profit. These findings provide strategic insights for automotive companies in formulating more effective, adaptive, and data-driven sales strategies.

Keywords: Business Intelligence; Quantitative Analysis; Random Forest Algorithm; Car Sales Analysis; Sales Performance

1. PENDAHULUAN

Industri otomotif di Amerika Serikat dalam beberapa tahun terakhir menghadapi perubahan signifikan dalam perilaku konsumen, termasuk meningkatnya minat terhadap kendaraan listrik yang memengaruhi keputusan pembelian dan struktur pasar mobil secara keseluruhan. Studi terbaru menunjukkan bahwa faktor-faktor seperti persepsi harga, kemudahan penggunaan teknologi, dan preferensi konsumen memainkan peran penting dalam menentukan keputusan untuk membeli kendaraan, khususnya kendaraan listrik dan termasuk pada pasar kendaraan listrik bekas, dimana mencerminkan kebutuhan pelaku industri untuk mengembangkan strategi penjualan yang lebih adaptif dan berbasis data guna mengidentifikasi faktor-faktor utama yang memengaruhi profit penjualan mobil [1],[2].

Sejalan dengan kondisi tersebut, berbagai penelitian akademik menunjukkan bahwa profitabilitas perusahaan otomotif dipengaruhi oleh sejumlah faktor internal, seperti likuiditas, leverage, dan ukuran perusahaan, yang terbukti memiliki pengaruh signifikan terhadap kinerja keuangan. Meskipun demikian, sebagian besar studi sebelumnya masih berfokus pada analisis keuangan konvensional dan belum banyak memanfaatkan pendekatan *machine learning* untuk mengeksplorasi faktor penentu profit secara lebih komprehensif pada level transaksi penjualan [3].

Dalam konteks penelitian ini, digunakan *dataset USA car sales* yang mencakup periode 2018–2024 dan memuat beragam variabel, antara lain harga jual, biaya, diskon, spesifikasi kendaraan, demografi pembeli, metode pembayaran, wilayah penjualan, serta komisi penjualan. Karakteristik data yang bersifat multi variabel dan kompleks menjadikan algoritma *random forest* sebagai metode yang relevan, karena kemampuannya dalam menangani data nonlinier, mengelola interaksi antar variabel, serta mengidentifikasi fitur-fitur yang paling berpengaruh melalui analisis *feature importance* [4],[5].

Penelitian terkait analisis profitabilitas penjualan mobil telah dilakukan oleh beberapa peneliti sebelumnya dengan pendekatan yang beragam. Penelitian oleh beberapa studi terdahulu menunjukkan bahwa analisis profit penjualan kendaraan umumnya dilakukan menggunakan metode statistik konvensional seperti regresi linier atau analisis keuangan untuk mengidentifikasi hubungan antara variabel harga, biaya operasional, dan kinerja penjualan. Penelitian lain memanfaatkan teknik data mining seperti *decision tree* untuk memprediksi kinerja penjualan kendaraan berdasarkan karakteristik pasar dan konsumen [6]. Selain itu, beberapa studi juga telah menerapkan algoritma *machine learning* seperti



support vector machine dan neural network untuk memprediksi tren penjualan kendaraan, namun sebagian besar penelitian tersebut lebih berfokus pada prediksi volume penjualan dibandingkan analisis faktor yang memengaruhi profit secara langsung [7]. Di sisi lain, masih terdapat keterbatasan penelitian yang secara khusus memanfaatkan algoritma ensemble learning seperti random forest untuk mengidentifikasi faktor-faktor paling berpengaruh terhadap profit penjualan mobil menggunakan data transaksi yang bersifat multivariabel [8],[9]. Oleh karena itu, penelitian ini berupaya mengisi celah penelitian tersebut dengan menerapkan algoritma random forest pada dataset transaksi penjualan mobil untuk mengidentifikasi faktor-faktor utama yang memengaruhi profit secara lebih komprehensif melalui analisis feature importance. Kontribusi penelitian ini terletak pada penerapan pendekatan machine learning berbasis ensemble untuk analisis profitabilitas penjualan kendaraan serta penyediaan pemahaman yang lebih mendalam mengenai variabel-variabel kunci yang memengaruhi profit sehingga dapat mendukung pengambilan keputusan bisnis yang lebih berbasis data dalam industri otomotif [10],[11].

Pemanfaatan analisis data berbasis machine learning dalam penelitian ini memungkinkan proses pengolahan, pemodelan, dan interpretasi data dilakukan secara lebih komprehensif. Dengan pendekatan tersebut, penelitian ini bertujuan untuk mengidentifikasi faktor-faktor utama yang memengaruhi profit penjualan mobil, sehingga hasilnya diharapkan dapat menjadi dasar yang kuat dalam mendukung pengambilan keputusan bisnis yang lebih efektif, strategis, dan berbasis data di industri otomotif.

2. METODOLOGI PENELITIAN

2.1 Jenis dan Sumber Data

Jenis data yang digunakan adalah data sekunder yaitu data yang dikumpulkan dari data yang telah ada sebelumnya, seperti data yang telah digunakan oleh peneliti sebelumnya atau dikumpulkan oleh organisasi atau instansi terkait. Data yang digunakan dalam penelitian ini berupa data sekunder yang diperoleh dari sumber Kaggle dengan judul dataset USA Car Sales 2018–2024 (yang dapat diakses melalui <https://www.kaggle.com/datasets/anjaliaprajapati307/usa-car-sales-dataset-2018-2024>).

2.2 Random Forest

Random Forest adalah salah satu algoritma dalam bidang *Machine Learning* yang digunakan untuk tugas klasifikasi dan regresi dengan membangun banyak pohon keputusan (decision tree) secara acak, kemudian menggabungkan hasil prediksi dari seluruh pohon tersebut untuk menghasilkan keputusan yang lebih akurat dan stabil [12],[13],[14],[15]. Algoritma ini bekerja dengan membuat beberapa *decision tree* dari subset data dan subset fitur yang dipilih secara acak (*random sampling*) [16],[17]. Setiap pohon memberikan prediksi, kemudian hasil akhir ditentukan melalui proses voting mayoritas pada klasifikasi atau rata-rata nilai pada regresi. Pendekatan ini merupakan bagian dari teknik *ensemble learning* yang bertujuan meningkatkan performa model serta mengurangi risiko *overfitting* [18],[19],[20].

2.3 Tahapan Penelitian

Berikut Tahapan penelitian yang digunakan penulis

a. Pengumpulan Data

Studi kepustakaan merupakan teknik pengumpulan data yang digunakan dalam penelitian ini melalui penelaahan terhadap buku, literatur ilmiah, serta jurnal yang relevan dengan permasalahan yang dikaji. Proses studi kepustakaan dilakukan untuk menghimpun data dan informasi teoretis yang digunakan sebagai dasar pendukung dalam penulisan, sehingga setiap argumen dan analisis yang disajikan dalam penelitian ini didasarkan pada sumber yang valid dan dapat dipertanggungjawabkan secara ilmiah. Data yang diperoleh sebagai penunjang penulisan jurnal berperan sebagai fondasi utama dalam penerapan metode *machine learning* (khususnya pada implementasi algoritma *random forest*), di mana kualitas dan kelengkapan data yang dikumpulkan sangat memengaruhi kinerja dan akurasi model. Oleh karena itu, pengumpulan data dalam jumlah yang memadai dan relevan menjadi aspek penting untuk memastikan model mampu mempelajari pola secara optimal dan menghasilkan hasil analisis yang andal.

b. Pemrosesan dan Pembersihan Data

Pemrosesan dan pembersihan data (*data preprocessing and cleaning*) merupakan langkah awal yang krusial dalam rangkaian *machine learning*, karena data mentah yang diperoleh biasanya mengandung banyak *noise*, inkonsistensi, ada kemungkinan beberapa nilai yang hilang, serta format yang tidak seragam yang dapat mengganggu kinerja model. Proses ini meliputi identifikasi dan penanganan nilai yang hilang (*missing values*), koreksi kesalahan entri, penghapusan duplikasi, serta penyesuaian format data agar setiap variabel berada dalam kondisi yang sesuai untuk dianalisis oleh algoritma. Data yang berkualitas baik merupakan fondasi utama dalam pemodelan *machine learning*, karena kualitas *input* akan sangat memengaruhi kemampuan model dalam belajar pola dan membuat prediksi yang akurat, tanpa tahapan *preprocessing* yang tepat, model dapat menghasilkan hasil prediksi yang bias, kurang stabil, atau bahkan tidak dapat berkonvergensi dengan benar.

c. Pembagian Data

Merupakan tahapan fundamental dalam pengembangan model *machine learning* yang bertujuan untuk memastikan proses evaluasi dilakukan secara objektif dan mencerminkan kemampuan model dalam menghadapi data baru. Secara umum, dataset dibagi menjadi dua subset utama, yaitu *training set* dan *testing set*, meskipun pada praktik tertentu



dapat digunakan lebih dari dua subset. *Training set* adalah bagian data yang digunakan untuk melatih model, di mana algoritma mempelajari pola, hubungan antar variabel, serta struktur data guna membentuk model prediksi atau klasifikasi. Sementara itu, *testing set* merupakan data yang dipisahkan dan tidak digunakan selama proses pelatihan, melainkan dimanfaatkan untuk mengukur kinerja model setelah pelatihan selesai

d. Pemilihan dan Penggunaan Algoritma

Pemilihan algoritma *machine learning* merupakan bagian penting dari pengembangan model karena keputusan ini akan memengaruhi kemampuan model dalam menangkap pola dan hubungan dalam data yang kompleks. Faktor-faktor yang mempengaruhi pemilihan algoritma meliputi tipe dan struktur data, tujuan pemodelan (misalnya klasifikasi atau regresi), serta ketersediaan sumber daya komputasi untuk pelatihan model. Dalam penelitian ini, algoritma yang dipilih adalah *random forest*, yang merupakan bagian dari *supervised learning*, yaitu kategori algoritma *machine learning* yang bekerja dengan data berlabel atau yang memiliki *output* yang diketahui.

e. Evaluasi Model

Merupakan tahapan penting dalam pengembangan sistem *machine learning* yang bertujuan untuk menilai sejauh mana model dapat menghasilkan prediksi yang akurat dan dapat digeneralisasi pada data yang tidak pernah dilihat selama proses pelatihan.

f. Feature Importance Analysis

Feature importance analysis dilakukan untuk mengidentifikasi variabel-variabel yang memiliki pengaruh paling signifikan terhadap hasil prediksi model *machine learning*. Dalam konteks algoritma *random forest*, *feature importance* diukur berdasarkan kontribusi setiap fitur terhadap pengurangan ketidakmurnian (*impurity*) di seluruh *decision tree* yang membentuk model, di mana fitur yang sering digunakan untuk memisahkan *node* pada pohon akan menunjukkan kontribusi yang lebih besar terhadap prediksi akhir model. Pendekatan ini memungkinkan peneliti untuk mendapatkan pemahaman kuantitatif mengenai variabel-variabel mana yang memberikan pengaruh terbesar dalam memprediksi profit, sehingga interpretasi hasil model menjadi lebih informatif dan dapat digunakan sebagai dasar penentuan prioritas faktor dalam pengambilan keputusan bisnis.

g. Visualisasi Dashboard

Digunakan untuk menyajikan hasil analisis data serta kinerja model dalam bentuk grafik dan visual interaktif yang mudah dipahami. *Dashboard* ini berfungsi untuk mempermudah interpretasi pola, tren, dan hubungan antar variabel yang memengaruhi profit penjualan mobil. Selain itu, dapat membantu menyampaikan temuan penelitian secara lebih informatif dan intuitif, sehingga mendukung proses pengambilan keputusan berbasis data.

h. Pelaporan Hasil

Tahap akhir dalam metode pengembangan sistem yang bertujuan untuk menyajikan dan menginterpretasikan seluruh rangkaian proses analisis yang telah dilakukan. Tahap ini mencakup penyajian hasil pemrosesan data, pemodelan menggunakan algoritma *random forest*, evaluasi kinerja model, serta visualisasi hasil analisis dalam bentuk tabel, grafik, dan *dashboard*. Pelaporan hasil disusun secara sistematis untuk memastikan bahwa temuan penelitian dapat dipahami dengan baik, transparan, serta mampu mendukung penarikan kesimpulan dan rekomendasi yang relevan bagi pengambilan keputusan berbasis data.

3. HASIL DAN PEMBAHASAN

Pada bagian ini menyajikan keluaran dari penerapan algoritma *random forest* dalam menganalisis faktor-faktor yang memengaruhi profit penjualan mobil, sekaligus membahas interpretasi hasil pemodelan dan analisis variabel secara komprehensif. Hasil evaluasi kinerja model dan analisis faktor penentu profit dibahas untuk memberikan pemahaman yang lebih mendalam mengenai peran masing-masing variabel dalam pembentukan profit penjualan mobil serta implikasinya terhadap pengambilan keputusan berbasis data.

3.1 Data Penelitian

Data yang digunakan dalam penelitian ini memuat 22 variabel utama terkait transaksi penjualan mobil. Dataset ini menggambarkan aktivitas penjualan kendaraan di berbagai wilayah Amerika Serikat, termasuk informasi mengenai konsumen, tipe kendaraan, harga, biaya transaksi, hingga detail waktu penjualan. Dataset berisi data transaksi penjualan mobil selama periode 7 tahun, setiap baris merepresentasikan satu transaksi penjualan mobil, sehingga dataset ini bersifat transaksional. Informasi yang dicakup sangat komprehensif, meliputi dimensi waktu, karakteristik pelanggan, spesifikasi kendaraan, detail harga, metode pembayaran, serta informasi wilayah penjualan dengan tujuan untuk memberikan gambaran menyeluruh mengenai faktor-faktor yang memengaruhi penjualan dan profitabilitas pada industri otomotif di Amerika Serikat. Berikut rincian variabel dari dataset, yaitu:

- Date*: Tanggal transaksi penjualan dilakukan.
- Salesperson*: Merupakan tenaga penjual yang menangani transaksi, setiap *salesperson* memiliki kemampuan negosiasi dan performa penjualan berbeda yang dapat berdampak pada nilai profit.
- Customer Name*: Nama pelanggan yang melakukan pembelian.
- Customer Age*: Menunjukkan usia pelanggan saat melakukan pembelian, dapat mencerminkan preferensi atau daya beli tertentu sehingga berpotensi memengaruhi besar kecilnya profit.



- e. *Customer Gender*: Jenis kelamin pelanggan (*male/ female*), variabel ini digunakan untuk melihat apakah terdapat kecenderungan perbedaan kontribusi profit berdasarkan kelompok gender.
- f. *Car Make*: Merek kendaraan yang dijual, seperti Ford, Kia, BMW, Audi, dan lainnya. Setiap merek memiliki rentang harga dan margin keuntungan yang berbeda sehingga memengaruhi potensi profit.
- g. *Car Model*: Tipe/ model mobil yang dibeli. Model tertentu memiliki popularitas dan nilai penjualan yang beragam sehingga dapat memengaruhi profit.
- h. *Car Year*: Tahun produksi kendaraan. Tahun produksi berkaitan dengan umur kendaraan dan daya tarik pasar yang ikut memengaruhi nilai profit.
- i. *Quantity*: Jumlah unit kendaraan yang dibeli pada transaksi tersebut. Semakin banyak jumlah pembelian, semakin besar profit yang mungkin diperoleh.
- j. *Sale Price*: Harga jual total kendaraan kepada pelanggan (sebelum diskon dan biaya lainnya).
- k. *Cost*: Biaya dasar/ biaya modal dari kendaraan yang menjadi dasar perhitungan profit.
- l. *Profit*: Selisih antara harga jual dan biaya ($Sale Price - Cost$), merupakan variabel utama yang dianalisis.
- m. *Discount*: Persentase diskon yang diberikan kepada pembeli. Diskon yang lebih besar berpotensi menurunkan profit sehingga menjadi variabel penting untuk dianalisis.
- n. *Payment Method*: Metode pembayaran yang digunakan (*Cash, Loan, Lease, dll.*). Preferensi metode pembayaran dapat menunjukkan karakteristik transaksi yang memengaruhi profit.
- o. *Commission Rate*: Persentase komisi yang diberikan kepada tenaga penjual.
- p. *Commission Earned*: Nominal komisi yang diterima tenaga penjual berdasarkan transaksi.
- q. *Sales Region*: Wilayah atau negara bagian tempat penjualan dilakukan. Setiap wilayah memiliki kondisi pasar, tingkat daya beli, dan pola permintaan yang berbeda sehingga dapat berdampak pada profit penjualan.
- r. *Sale Year*: Tahun terjadinya penjualan. Digunakan untuk mengamati tren profit dari waktu ke waktu selama tahun 2018–2024.
- s. *Sale Month*: Bulan transaksi (*January–December*). Variabel ini digunakan untuk melihat pola musiman penjualan yang mungkin memengaruhi nilai profit.
- t. *Sale Quarter*: Kuartal transaksi (Q1–Q4). Kuartal tertentu biasanya memiliki performa penjualan lebih tinggi atau lebih rendah sehingga mempengaruhi profit.
- u. *Day of Week*: Hari terjadinya transaksi (*Monday–Sunday*). Penjualan kemungkinan lebih tinggi pada hari-hari tertentu sehingga berdampak pada profit.
- v. *Season*: Musim pada waktu transaksi (*Winter, Spring, Summer, Fall*). Musim dapat memengaruhi perilaku pembelian mobil dan akhirnya memengaruhi profit.

Berdasarkan perumusan masalah dan tujuan penelitian, model yang dibangun dalam penelitian ini melibatkan satu variabel dependen (target) dan sejumlah variabel independen (prediktor) dari 22 variabel yang di dapat dari dataset. Variabel-variabel tersebut dipilih untuk merepresentasikan karakteristik penjualan, harga, dan faktor operasional yang diduga berpengaruh terhadap profit penjualan mobil. Uraian berikut menjelaskan peran dan definisi masing-masing variabel yang digunakan dalam pemodelan.

a. Variabel Target

Variabel target merupakan variabel yang menjadi fokus utama dalam pemodelan dan berfungsi sebagai keluaran yang ingin diprediksi atau dijelaskan oleh model. Secara praktis, variabel target adalah respons yang nilainya bergantung pada berbagai *input* atau prediktor yang dimasukkan ke dalam model dan dijadikan tolok ukur utama keberhasilan prediksi, sehingga variabel ini umumnya merepresentasikan hasil akhir yang relevan dalam konteks analisis bisnis. Pada penelitian ini, variabel target yang digunakan adalah profit. Profit didefinisikan sebagai keuntungan bersih yang diperoleh dari setiap transaksi penjualan mobil. Nilai profit dipengaruhi oleh beberapa komponen, seperti harga jual, besaran diskon, serta jumlah unit mobil yang terjual dalam satu transaksi. Dalam penelitian ini, profit digunakan sebagai indikator utama untuk mengukur kinerja penjualan dan mengevaluasi sejauh mana variabel-variabel prediktor berkontribusi terhadap peningkatan atau penurunan keuntungan. Algoritma Random Forest diterapkan untuk menganalisis hubungan antara variabel prediktor dan nilai profit, serta mengidentifikasi tingkat pengaruh masing-masing faktor terhadap profit penjualan mobil.

b. Variabel Prediktor

Dalam konteks *machine learning* dan pemodelan prediktif, variabel independen (prediktor) merupakan variabel-variabel input yang digunakan oleh model untuk memperkirakan atau menjelaskan variasi dalam variabel dependen (target); variabel prediktor ini mencerminkan faktor-faktor yang diduga memengaruhi *outcome* yang diamati dan dapat berupa karakteristik demografis, atribut bisnis, spesifikasi produk, atau waktu dan konteks transaksi yang relevan dengan fenomena yang dimodelkan. Dalam suatu studi prediksi, seperti yang digunakan untuk memodelkan profit penjualan mobil, variabel prediktor memainkan peran sebagai sumbu informasi yang menjadi dasar bagi algoritma untuk “belajar” pola hubungan antara *input* dan target, sehingga perubahan pada prediktor dapat memberikan *insight* mengenai sejauh mana setiap faktor berkontribusi terhadap prediksi hasil tertentu. Secara umum, variabel prediktor dipilih berdasarkan relevansinya terhadap tujuan analisis dan ketersediaannya dalam dataset, dan perannya adalah sebagai *input* yang diolah oleh model untuk menghasilkan estimasi atau klasifikasi target variabel. Variabel prediktor dalam penelitian ini berjumlah 16, yaitu: *Customer age, customer gender, salesperson, car make, car model, car year, quantity, discount, payment method, sales region, sale year, sale month, sale quarter, day of week, season*, dan terakhir



adalah variabel *car age* (mempunyai rumus $sale\ year - car\ year$) yang merupakan usia mobil, semakin tua usia kendaraan, nilai jualnya cenderung lebih rendah sehingga berdampak pada profit.

3.2 Implementasi Algoritma *Random Forest*

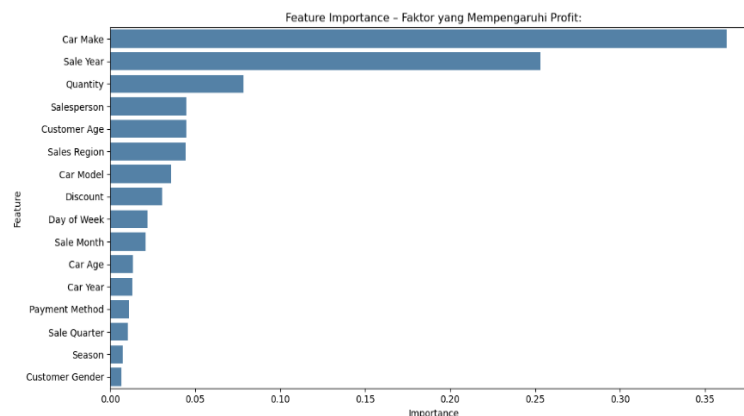
Berdasarkan tahapan metode pengembangan sistem yang telah dijelaskan sebelumnya, bagian selanjutnya menyajikan hasil implementasi dari model dan sistem yang dibangun. Bagian ini memaparkan bagaimana proses pemodelan dan analisis tersebut diwujudkan dalam bentuk implementasi serta hasil yang diperoleh dari penerapannya pada data penelitian.

Hasil perhitungan *feature importance* dari algoritma *random forest* yang digunakan untuk memprediksi nilai profit. Berdasarkan data yang dihasilkan tersebut, variabel *Car Make* memiliki nilai kepentingan tertinggi sebesar 0.362573, yang menunjukkan bahwa merek mobil merupakan faktor paling dominan dalam menentukan besarnya profit setiap transaksi. Variabel *Sale Year* berada pada posisi kedua dengan nilai 0.253027, mengindikasikan bahwa tren penjualan berdasarkan tahun memberikan pengaruh signifikan terhadap profitabilitas. Selanjutnya, *Quantity* menjadi faktor penting ketiga dengan nilai 0.078557, yang berarti jumlah unit mobil yang dibeli dalam satu transaksi turut berkontribusi nyata terhadap profit. Variabel lainnya seperti *Salesperson*, *Customer Age*, dan *Sales Region* memiliki nilai kepentingan lebih rendah, sementara faktor seperti *Season*, *Sale Quarter*, dan *Customer Gender* menunjukkan pengaruh minimal dalam model.

Feature	Importance
3 Car Make	0.362573
10 Sale Year	0.253027
6 Quantity	0.078557
2 Salesperson	0.044895
0 Customer Age	0.044777
9 Sales Region	0.044561
4 Car Model	0.036813
7 Discount	0.030498
13 Day of Week	0.022895
11 Sale Month	0.020763
15 Car Age	0.013422
5 Car Year	0.013024
8 Payment Method	0.011315
12 Sale Quarter	0.010549
14 Season	0.007355
1 Customer Gender	0.006577

Gambar 1. Hasil *feature importance*

Selanjutnya, pada Gambar 2 menampilkan tingkat pengaruh antar variabel terhadap profit, grafik ini menunjukkan dengan jelas bahwa *Car Make* menempati posisi paling atas dengan batang yang paling panjang, mempertegas bahwa pemilihan merek mobil merupakan faktor utama dalam menentukan nilai profit. Diikuti oleh *Sale Year*, grafik memperlihatkan bahwa perubahan tahun penjualan berperan signifikan dalam membentuk profit secara keseluruhan. *Quantity* juga tampak memberikan kontribusi penting, meskipun tidak sebesar dua variabel sebelumnya. Sementara itu, batang untuk variabel seperti *Payment Method*, *Sale Quarter*, *Season*, dan *Customer Gender* terlihat sangat pendek, menandakan bahwa variabel-variabel tersebut memiliki kontribusi yang relatif kecil dalam prediksi profit. Ini semua membantu dalam memberikan gambaran intuitif mengenai variabel mana yang paling berpengaruh dan mana yang memiliki peran terbatas dalam model.

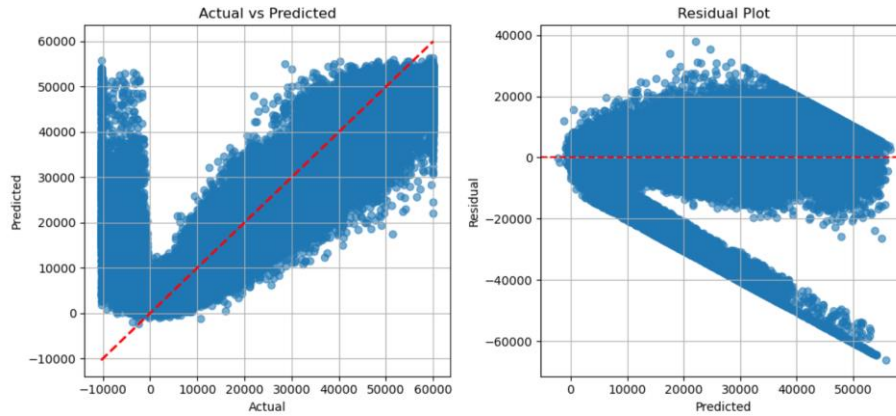


Gambar 2. Grafik *feature importance*

Evaluasi model dilakukan menggunakan nilai *R-squared* (R^2), *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE). Model menghasilkan nilai R^2 sebesar 0,6841, yang menunjukkan bahwa sekitar 68,41% variasi profit dapat dijelaskan oleh fitur-fitur yang digunakan dalam model. Nilai MAE sebesar 4.945,1 mengindikasikan tingkat deviasi prediksi profit per transaksi berada pada kisaran 4–5 ribu dolar. Hal ini menunjukkan bahwa model memiliki akurasi prediksi yang cukup baik, meskipun masih ditemukan deviasi pada transaksi dengan nilai profit ekstrem, sejalan dengan hasil RMSE yang lebih tinggi akibat sensitivitas terhadap *outlier*. Sementara itu, nilai sebesar 8020,05

mengindikasikan adanya sejumlah kesalahan prediksi yang lebih besar akibat sensitivitas metrik ini terhadap nilai ekstrem.

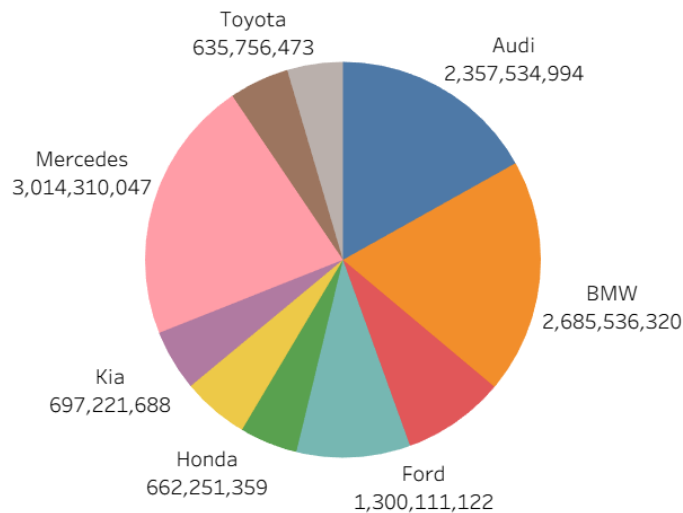
Gambar 3 melalui grafik *Actual vs Predicted* dan *Residual Plot*, menunjukkan bahwa sebagian besar titik berada di sekitar garis diagonal acuan, yang menandakan bahwa model mampu menghasilkan prediksi yang cukup mendekati nilai aktual. Namun, terlihat kecenderungan *overprediction* pada nilai profit rendah serta *underprediction* pada nilai profit tinggi. *Residual plot* menunjukkan penyebaran residual yang semakin melebar pada nilai prediksi besar, yang mengindikasikan adanya heteroskedastisitas serta variasi data yang belum sepenuhnya dapat ditangkap oleh model.



Gambar 3. Grafik Actual vs Predicted dan Residual Plot

3.3 Visualisasi

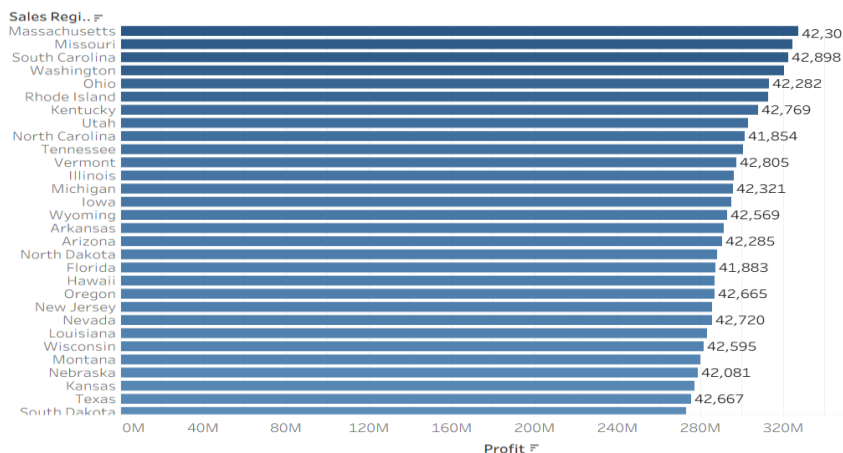
Visualisasi data menggunakan Tableau diterapkan untuk menyajikan hasil analisis secara lebih jelas, sistematis, dan informatif. Tableau merupakan perangkat lunak visualisasi data yang banyak digunakan dalam penelitian berbasis *business intelligence* untuk menampilkan temuan analisis dalam bentuk grafik dan dashboard interaktif. Dalam penelitian ini, Tableau dimanfaatkan untuk memperkuat interpretasi hasil pemodelan dengan menampilkan pola penjualan, tren profit, serta distribusi variabel-variabel penting yang memengaruhi profit penjualan mobil. Penyajian hasil analisis melalui visualisasi ini memudahkan pemahaman terhadap pola dan hubungan data, sehingga dapat mendukung proses interpretasi hasil serta pengambilan keputusan bisnis yang lebih berbasis data.



Gambar 4. Visualisasi profit per brand

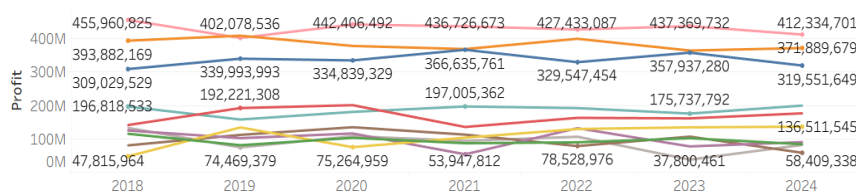
Dari visualisasi di atas terlihat bahwa Mercedes menghasilkan profit tertinggi dengan total lebih dari 3 triliun, diikuti oleh BMW dan Audi yang juga memberikan kontribusi signifikan. Sementara itu, merek seperti Toyota, Honda, dan Kia berada pada posisi profit yang lebih rendah dibandingkan merek premium lainnya. Hal ini menunjukkan bahwa kendaraan dari segmen premium memiliki *margin* keuntungan yang lebih besar dan memberikan dampak yang dominan terhadap total profit.

Sedangkan profit berdasarkan wilayah penjualan tercermin dalam Gambar 5. Massachusetts, Missouri, dan South Carolina muncul sebagai wilayah dengan profit tertinggi, sementara wilayah seperti South Dakota dan Kansas berada di posisi bawah. Perbedaan ini menunjukkan bahwa profit dipengaruhi oleh faktor regional seperti daya beli, *volume* transaksi, dan kondisi pasar di masing-masing negara bagian. Visualisasi ini membantu mengidentifikasi wilayah dengan kontribusi profit paling besar serta area yang memiliki potensi peningkatan strategi pemasaran.



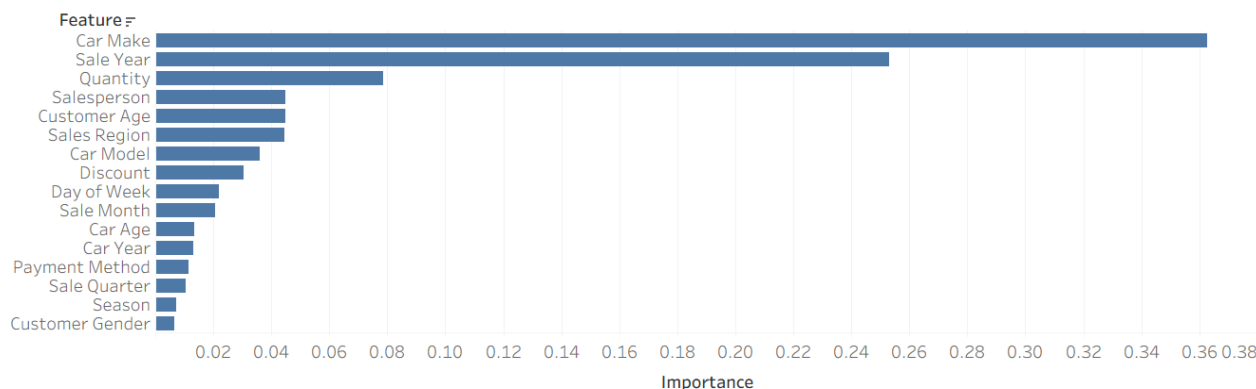
Gambar 5. Visualisasi profit per Sales Region

Gambar 6 menampilkan tren profit per tahun dari 2018 hingga 2024. Terlihat bahwa profit mengalami fluktuasi setiap tahun, dengan puncak profit terjadi pada tahun 2020 dan sedikit menurun pada tahun 2021–2024. Beberapa merek menunjukkan pola kenaikan dan penurunan yang berbeda-beda, mencerminkan variasi permintaan konsumen dan kondisi pasar pada periode tersebut. Visualisasi ini memberikan gambaran mengenai dinamika profit tahunan serta membantu memahami perubahan tren penjualan dari waktu ke waktu.



Gambar 6. Visualisasi profit per tahun

Analisis *feature importance* menggunakan Tableau yang ditampilkan pada Gambar 7 menunjukkan variabel-variabel yang paling berpengaruh dalam membentuk nilai profit berdasarkan hasil algoritma *random forest*. Dari grafik tersebut terlihat bahwa *Car Make* menjadi faktor dengan pengaruh terbesar, diikuti oleh *Sale Year* dan *Quantity*, yang menandakan bahwa merek mobil, tahun penjualan, serta jumlah unit terjual sangat menentukan profit. Variabel seperti *Salesperson*, *Customer Age*, dan *Sales Region* memiliki tingkat pengaruh menengah, sedangkan fitur seperti *Season*, *Sale Quarter*, dan *Customer Gender* menunjukkan kontribusi yang sangat kecil. Visualisasi ini membantu memperjelas prioritas variabel yang paling memengaruhi profit serta mendukung interpretasi hasil model secara lebih intuitif.



Gambar 7. Visualisasi hasil analisis *feature importance*

4. KESIMPULAN

Berdasarkan hasil analisis faktor-faktor yang memengaruhi profit penjualan mobil, dapat disimpulkan bahwa variabel *car make*, *sale year*, dan *quantity* merupakan faktor yang paling berpengaruh dalam menentukan besarnya profit. Ketiga variabel tersebut memiliki nilai *feature importance* tertinggi pada algoritma *random forest*, yang menunjukkan bahwa merek mobil, tahun terjadinya penjualan, serta jumlah unit yang dibeli memberikan kontribusi dominan terhadap pembentukan profit. Variabel lain seperti *salesperson*, *customer age*, dan *sales region* menunjukkan pengaruh pada tingkat menengah, sedangkan variabel *season*, *sale quarter*, dan *customer gender* memiliki pengaruh yang relatif kecil.



terhadap profit penjualan. Hasil evaluasi model menunjukkan bahwa algoritma *random forest* mampu menjelaskan variasi profit dengan nilai R^2 sebesar 0,6841, yang berarti sekitar 68,41% variasi profit dapat dijelaskan oleh model. Temuan ini mengindikasikan bahwa model memiliki kemampuan prediktif yang kuat dalam mengidentifikasi kontribusi variabel-variabel penentu profit. Selain itu, pemanfaatan visualisasi berbasis *business intelligence* melalui Tableau memperkuat hasil analisis dengan menampilkan pola profit berdasarkan merek, wilayah, dan tahun penjualan secara informatif. Secara keseluruhan, penelitian ini menegaskan bahwa integrasi pendekatan *machine learning* dan *business intelligence* efektif dalam mengidentifikasi faktor penentu profit penjualan mobil serta dapat menjadi dasar yang andal dalam mendukung pengambilan keputusan bisnis yang lebih strategis dan berbasis data.

REFERENCES

- [1] S. T. Hamidou and A. Mehdi, "Machine Learning with Applications Enhancing IDS performance through a comparative analysis of Random Forest, XGBoost, and Deep Neural Networks," *Mach. Learn. with Appl.*, vol. 22, no. September, p. 100738, 2025, doi: 10.1016/j.mlwa.2025.100738.
- [2] P. Aulia Azhar, M. Arya Pratama, and R. Fitriani, "Prediksi Harga Mobil Audi Bekas Menggunakan Model Regresi Linear dengan Framework Streamlit," *J. Technol. Informatics*, vol. 6, no. 1, pp. 22–28, 2024, doi: 10.37802/joti.v6i1.763.
- [3] W. Wilianto, Y. Yuliana, A. Suwandhi, J. Jimmy, and J. Putra, "Penerapan AI dalam Menentukan Harga Mobil Bekas Berdasarkan Tahun Perakitan," *J. Minfo Polgan*, vol. 13, no. 1, pp. 550–560, 2024, doi: 10.33395/jmp.v13i1.13728.
- [4] A. Prayoga, Y. V. Via, and I. DIYASA, "Classifying Legendary Pokémon with SF-Random Forest Algorithm," *J. Inf. Syst. Informatics*, vol. 6, no. 3, pp. 1852–1871, 2024.
- [5] R. Hidayat *et al.*, "Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket," *Simkom*, vol. 10, no. 1, pp. 101–109, 2025, doi: 10.51717/simkom.v10i1.703.
- [6] Y. Liguori, I. W. Sudiarsa, I. M. J. Dita, I. G. N. Galih, and J. Baskara, "Implementasi Algoritma Random Forest untuk Klasifikasi Rentang Harga Ponsel Berdasarkan Spesifikasi Teknis," *J. ABDINUS J. Pengabd. Nasant.*, no. November, 2025.
- [7] Y. Sutanto, B. Al Amin, H. A. Setyadi, and B. E. Purnama, "Prediksi Harga Perumahan Menggunakan Metode Principal Housing Price Prediction Using Principal Component Analysis and Random Forest Regression Methods," *J. Inform.*, vol. 12, no. 6, pp. 1243–1250, 2025.
- [8] A. A. Karim, M. A. Prasetyo, and M. R. Saputro, "Perbandingan Metode Random Forest, K-Nearest Neighbor, dan SVM Dalam Prediksi Akurasi Pertandingan Liga Italia," *Pros. Semin. Nas. Teknol. dan Sains*, vol. 2, no. 3, pp. 377–342, 2023.
- [9] S. A. S. Mola, Y. C. Luttu, and D. N. Rumlaklak, "Perbandingan Metode Machine Learning dalam Analisis Sentimen Komentar Pengguna Aplikasi InDriver pada Dataset Tidak Seimbang," *J. Sist. Inf. Bisnis*, vol. 14, no. 3, pp. 247–255, 2024, doi: 10.21456/vol14iss3pp247-255.
- [10] A. R. Masdian, N. Bashit, and F. Hadi, "Analisis Produktivitas Padi Menggunakan Algoritma Machine Learning Random Forest Di Kabupaten Batang Tahun 2018-2022," *Elipsoida J. Geod. dan Geomatika*, vol. 6, no. 1, pp. 43–51, 2023.
- [11] I. Kurniawan, D. C. P. Buani, A. Abdussomad, W. Apriliah, and R. A. Saputra, "Implementasi algoritma random forest untuk menentukan penerima bantuan raskin," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 10, no. 2, pp. 421–428, 2023.
- [12] R. Hanifudin, P. Rokhmayati, M. Fadhly Noor Rizqi, and L. Fitriana Masitoh, "Rancang Bangun Sistem Enterprise Resource Planning (ERP) Berbasis Web pada Pt Sainsgo Karya Indonesia Menggunakan Metode Scrum," *Syntax Idea*, vol. 6, no. 6, pp. 2857–2871, 2024, doi: 10.46799/syntax-idea.v6i6.3889.
- [13] J. Yandi and K. Wijaya, "Rancang Bangun Aplikasi Persediaan Barang Pada Counter Karya Cellmenggunakan Vb Net," *JSK (Jurnal Sist. Inf. dan Komputerisasi Akuntansi)*, vol. 7, no. 1, pp. 1–6, 2023, doi: 10.56291/jsk.v7i1.109.
- [14] S. Sobari, A. I. Purnamasari, A. Bahtiar, and K. Kaslani, "Meningkatkan Model Prediksi Kelulusan Santri Tahfidz di Pondok Pesantren Al-Kautsar Menggunakan Algoritma Random Forest," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, 2025.
- [15] I. L. Mulyahati, "Implementasi machine learning prediksi harga sewa apartemen menggunakan algoritma Random Forest melalui framework website Flask Python pada website mamikos.com," *Data Sci.*, vol. 1, no. 3, pp. 371–379, 2020.
- [16] R. P. Munggaran and M. Nurmalasari, "Predicting Outpatient Service Waiting Times with Random Forest Algorithm," *Data Sci. Indones.*, vol. 5, no. 1, pp. 35–40, 2025.
- [17] R. Faizal, A. Abdullah, and M. W. Pangestika, "Perbandingan Random Forest Regressor Dan Decision Tree Regressor Untuk Prediksi Hasil Panen," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 6, no. 2, pp. 247–253, 2025.
- [18] H. Sunaryanto, M. A. Hasan, and G. Guntoro, "Classification Analysis of Unilak Informatics Engineering Students Using Support Vector Machine (SVM), Iterative Dichotomiser 3 (ID3), Random Forest and K-Nearest Neighbors (KNN)," *IT J. Res. Dev.*, vol. 7, no. 1, pp. 36–42, 2022, doi: 10.25299/itjrd.2022.8912.
- [19] E. Fitri and S. N. Nugraha, "Optimasi Kinerja Linear Regression, Random Forest Regression Dan Multilayer Perceptron Pada Prediksi Hasil Panen," *Inti Nusa Mandiri*, vol. 18, no. 2, pp. 210–217, 2024.
- [20] N. Maulidah, M. Maulidah, R. Supriyadi, H. Nalatissifa, S. Diantika, and A. Fauzi, "Prediksi Kualitas Air Menggunakan Metode Random Forest, Decision Tree, Dan Gradient Boosting," *J. Khatulistiwa Inform.*, vol. 12, no. 1, pp. 1–6, 2024, doi: 10.31294/jki.v12i1.16004.