



INNOVATIVE: Journal Of Social Science Research

Volume 4 Nomor 4 Tahun 2024 Page 1199-1216

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Prediksi Penjualan Kendaraan Menggunakan *Regresi Linear*: Studi Kasus pada Industri Otomotif di Indonesia

Ilham Amansyah^{1✉}, Jamaludin Indra², Euis Nurlaelasari³, Ayu Ratna Juwita⁴

Universitas Buana Perjuangan Karawang

Email: if20.ilhamamansyah@mhs.ubpkarawang.ac.id^{1✉}

Abstrak

Industri otomotif Indonesia memiliki tingkat persaingan yang tinggi, sehingga perusahaan mobil seperti Toyota membutuhkan prediksi penjualan yang akurat untuk perencanaan bisnis yang efektif, dan prediksi penjualan yang akurat sangat penting untuk perencanaan bisnis yang efektif. Penelitian ini bertujuan untuk mengaplikasikan algoritma Regresi Linear dalam meramalkan penjualan mobil Toyota di Negara Indonesia. Data penjualan yang digunakan dalam penelitian ini diperoleh dari laporan penjualan mobil Toyota periode 2018 hingga 2023 yang diterbitkan oleh Gabungan Industri Kendaraan Bermotor Indonesia (GAIKINDO). Penelitian ini meliputi beberapa tahapan, mulai dari analisis masalah, pengumpulan data, preprocessing data, penerapan algoritma regresi linier, hingga evaluasi model menggunakan mean absolute error (MAE), mean square error (MSE), square error average (RMSE). dan rata-rata persentase kesalahan absolut (MAPE). Hasil penelitian menunjukkan bahwa model regresi linier dapat memprediksi penjualan mobil Toyota dengan akurasi yang cukup baik, dengan rata-rata kesalahan mutlak (MAE) sebesar 2.617 Unit penjualan dan rata-rata persentase kesalahan absolut (MAPE) sebesar 12,47% yang menunjukkan tingkat yang baik dalam akurasi ramalan. Nilai MAE, MSE, RMSE, Mape yang rendah menunjukkan bahwa model ini efektif dalam meramalkan penjualan di masa depan. Prediksi penjualan mobil Toyota untuk beberapa bulan ke depan menunjukkan hasil yang mendekati nilai aktual, sehingga model ini dapat diandalkan untuk perencanaan bisnis yang lebih baik.

Kata Kunci: *Algoritma Regresi Linear, Prediksi Penjualan, Industri Otomotif, Data Mining, Tren Penjualan*

Abstract

The Indonesian automotive industry has a high level of competition, so car companies like Toyota need accurate sales predictions for effective business planning, and accurate sales predictions are very important for effective business planning. This research aims to apply the Linear Regression algorithm in predicting Toyota car sales in Indonesia. The sales data used in this research was obtained from Toyota car sales reports for the period 2018 to 2023 published by the Association of Indonesian Automotive Industries (GAIKINDO). This research includes several stages, starting from problem analysis, data collection, data preprocessing, application of linear regression algorithms, to model evaluation using mean absolute error (MAE), mean square error (MSE), average square error (RMSE), and mean absolute percentage error (MAPE). The results of the research show that the linear regression model can predict Toyota car sales with fairly good accuracy, with an average absolute error (MAE) of 2,617 sales units and an average percentage absolute error (MAPE) of 12.47% which shows a high level of accuracy. good in forecast accuracy. The low MAE, MSE, RMSE, Mape values indicate that this model is effective in predicting future sales. Toyota car sales predictions for the next few months show results that are close to actual values, so this model can be relied on for better business planning.

Keywords: *Linear Regression Algorithms, Sales Prediction, Automotive Industry, Data Mining, Sales Trends*

PENDAHULUAN

Industri kendaraan memiliki peran penting dalam perekonomian global yang terus berkembang, untuk mengoptimalkan operasi bisnis yang berasal dari kebutuhan ekonomi Masyarakat. Toyota adalah salah satu produsen mobil terbesar di dunia saat ini dan telah lama berperan dalam industri otomotif global. Perusahaan ini dikenal dengan berbagai varian mobil, mulai dari kendaraan keluarga hingga kendaraan bisnis.

Produsen dan distributor kendaraan memerlukan prediksi akurat mengenai penjualan di masa depan. Dalam konteks perencanaan, membuat prediksi atau ramalan penjualan menjadi aspek krusial, yang menjadi landasan penting sebelum merencanakan suatu kegiatan atau proyek. Seiring perkembangan waktu, jumlah merek mobil pun semakin bertambah.

Dalam analitik modern, penggunaan algoritma data mining dan machine learning telah menjadi metode yang umum untuk peramalan bisnis, yang melibatkan prediksi perkembangan bisnis di masa depan berdasarkan analisis tren data historis dan data

terkini. Penggunaan teknik data mining untuk meramalkan penjualan memberikan kemampuan untuk memperkirakan jumlah penjualan kendaraan dalam beberapa bulan mendatang.

Berbagai algoritma prediksi, termasuk *Regresi Linear*, *Regresi Logistik*, dan *Decision Tree*. Fokus utama penelitian ini ditempatkan pada algoritma *Regresi Linear*, yang terkenal dan umum digunakan. Metode *Regresi Linear* digunakan untuk memodelkan dan memprediksi penjualan kendaraan roda empat. Hasil pengujian keakuratan menunjukkan bahwa prediksi penjualan kendaraan dengan menggunakan *Regresi Linear* dalam penelitian ini dinilai sangat baik.

Dengan menerapkan analisis data mining yang telah dikembangkan, masalah yang selama ini muncul dalam proses perhitungan prediksi peningkatan omset dapat diatasi. Proses perhitungan menjadi lebih sederhana, menghemat waktu, dan menghasilkan data yang lebih akurat untuk pengambilan keputusan.

Ketidakpastian dalam meramalkan penjualan kendaraan dapat mengakibatkan perencanaan bisnis yang tidak efektif, seperti produksi yang tidak sesuai dengan permintaan aktual. Prediksi yang akurat membantu menghindari kelebihan stok atau kekurangan persediaan. Oleh karena itu, tujuan penelitian ini adalah membangun model *Regresi Linear* yang dapat memprediksi penjualan kendaraan dengan tingkat akurasi yang tinggi, dengan memperhitungkan variabilitas pasar dan faktor-faktor eksternal yang memengaruhi penjualan.

Langkah berikutnya adalah menerapkan algoritma *Regresi Linear* dalam proses Data Mining untuk mengolah dataset. Pemanfaatan data mining bertujuan untuk mengelola dataset besar, mendukung penyimpanan, dan pengolahan data penjualan guna memperoleh informasi yang diperlukan. Dalam konteks prediksi, data mining digunakan secara sistematis untuk memperkirakan kejadian di masa depan berdasarkan informasi masa lalu dan saat ini. Tujuannya adalah mengurangi tingkat kesalahan dalam perkiraan yang dibuat.

Beberapa penelitian telah menunjukkan keberhasilan berbagai algoritma prediksi dalam konteks yang berbeda. Sebagai contoh, *Regresi Logistik* dapat memprediksi keberhasilan pemasaran jasa perbankan dengan akurasi 92,32%, sementara *Decision Tree* dapat memprediksi persediaan obat dan penjualan di apotek dengan tingkat akurasi 89%. Penggunaan *Regresi Linear* berganda pada SPSS24 untuk memproyeksikan penjualan mobil di tahun mendatang juga menunjukkan tingkat kesalahan estimasi relatif kecil, sekitar 0,0607, menegaskan kemampuan *Regresi Linear* dalam memprediksi penjualan

mobil secara akurat. Dalam konteks harga jual mobil bekas, penelitian menggunakan *Regresi Linear* berhasil mencapai tingkat akurasi sebesar 80%, menunjukkan kecukupan metode tersebut dalam kasus tersebut.

Dari hasil beberapa penelitian yang telah dilakukan, Algoritma *Regresi Linear* menunjukkan tingkat akurasi yang tinggi dan dapat menghasilkan hasil prediksi yang berbeda dari penelitian-penelitian sebelumnya. Penggunaan metode ini menjadi efektif karena mampu memahami hubungan linier antara variabel *dependen* dan lebih dari satu variabel *independen*. Algoritma *Regresi Linear* terbukti efisien dalam peramalan, khususnya dalam situasi yang melibatkan lebih dari dua faktor yang memengaruhi variabel *dependen*. Metode ini menghasilkan hasil terbaik dalam memahami hubungan kompleks antara variabel *independen* dan variabel *dependen*.

Menurut data dari Gabungan Industri Kendaraan Bermotor Indonesia (GAIKINDO), pada tahun 2018, penjualan mobil di Indonesia mencapai 1,1 juta unit per tahun. Tingginya potensi pasar di Indonesia ini menyebabkan persaingan yang sengit di antara perusahaan-perusahaan dalam industri ini.

Penelitian ini bertujuan untuk meneliti penggunaan Algoritma *Regresi Linear* dalam meramalkan penjualan kendaraan mobil Toyota di industri otomotif Indonesia. Dalam proyek ini, akan diperkirakan penjualan mobil Toyota untuk 6 bulan ke depan. Namun, tidak dimiliki data penjualan aktual untuk membandingkan hasil prediksi yang diperoleh sehingga tingkat kesuksesan prediksi tidak dapat dilihat. Meskipun demikian, analisis *Regresi Linear* diharapkan dapat memberikan gambaran perkiraan yang berguna dalam mengidentifikasi tren dan pola penjualan potensial.

METODE PENELITIAN

Penelitian ini dimulai dengan melakukan tinjauan literatur, yang bertujuan untuk menemukan dasar teoritis yang akan digunakan serta mencari literatur ilmiah yang relevan sebagai dukungan bagi penelitian ini. Selama keseluruhan proses penelitian, langkah-langkah atau tahapan penelitian dapat dijelaskan sebagai berikut :



Gambar 1. Tahapan Penelitian

Pada Gambar 1. Tahapan Penelitian diatas adalah alur dari metode penelitian yang dibuat, dimulai dari Analisis masalah, Pengumpulan data, Memproses data, Penerapan, Hasil dan pemodelan, sampai evaluasi Algoritma *Regresi Linear*.

Pada Gambar 1. Tahapan Penelitian diatas adalah alur dari metode penelitian yang dibuat, dimulai dari Analisis masalah, Pengumpulan data, Memproses data, Penerapan, Hasil dan pemodelan, sampai evaluasi Algoritma *Regresi Linear*.

Analisis Masalah

Industri otomotif di Indonesia menghadapi masalah utama terkait tingkat ketidakpastian tinggi dalam prediksi penjualan kendaraan. Prediksi yang tidak akurat dapat menyebabkan kesalahan dalam perencanaan produksi, tidak seimbangan persediaan, dan dampak finansial yang signifikan bagi perusahaan otomotif. Faktor-faktor seperti fluktuasi suku bunga, pertumbuhan ekonomi yang tidak stabil, dan perubahan kebijakan pajak menambah kompleksitas prediksi pasar. Metode prediksi saat ini mungkin tidak optimal dalam menghadapi variasi pasar yang cepat dan dinamis. Keterbatasan dalam model prediksi dapat mengakibatkan estimasi yang tidak akurat, dapat menghambat respons yang tepat terhadap perubahan kondisi pasar.

Oleh karena itu, diperlukan pendekatan yang lebih canggih dan diperbarui untuk meningkatkan akurasi prediksi penjualan kendaraan. Kurangnya pemahaman mendalam tentang faktor-faktor eksternal yang memengaruhi penjualan, seperti perubahan regulasi pajak atau kebijakan pemerintah, juga menjadi hambatan yang perlu diatasi melalui analisis mendalam untuk mengurangi risiko ketidakpastian.

Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah data penjualan mobil Toyota dari tahun 2018 hingga 2023. Data tersebut berasal dari laporan yang dikeluarkan oleh Gabungan Industri Kendaraan Bermotor Indonesia (GAIKINDO). [21].

Tabel 1. Data Penjualan Mobil Toyota Tahun 2018-2023

No.	Tahun - Bulan	Penjualan
1	2018-01	25405
2	2018-02	27665
3	2018-03	31424
4	2018-04	29360
5	2018-05	28950
...
...
...
68	2023-08	28104
69	2023-09	26600
70	2023-10	28003
71	2023-11	29125
72	2023-12	30825

Dalam tabel tersebut terdapat variabel x Ini adalah variabel kategorikal yang menunjukkan periode waktu penjualan mobil. Variabel ini terdiri dari dua bagian: tahun dan bulan. Gabungan keduanya menjadi variabel independen yang menentukan variabel dependen (penjualan). Sedangkan variabel y variabel numerik yang menunjukkan jumlah mobil Toyota yang terjual dalam periode waktu tertentu. Variabel ini merupakan variabel dependen yang dipengaruhi oleh variabel independen (tahun - bulan).

Preprocessing

Sebelum menerapkan *Regresi Linear*, penting untuk melakukan tahap *preprocessing data* untuk memastikan kualitas dan ketepatan model. Terdapat tahap *Data Cleaning* dimana data yang diperoleh akan diperiksa untuk memastikan bahwa data tidak terdapat nilai yang hilang dan duplikat data.

Dataset diperiksa menyeluruh untuk memastikan dan mendeteksi nilai yang hilang dan duplikat, yaitu dengan mencetak jumlah nilai yang hilang, mencetak bentuk *Data Frame*, dan menghitung jumlah duplikat. Berikut adalah langkah-langkah yang dapat

diambil untuk mengatasi masalah *Preprocessing pada tahap Data Cleaning*, disertakan dengan *Pseudocode* :

```
Mulai
Mencari Nilai Hilang(arr: Array of Integer)
  missingValue = -1
  found = False

  for i from 1 to arr.length
    if arr[i] != i
      missingValue = i
      found = True
      exit loop
    end if
  end for

  if not found
    missingValue = arr.length + 1
  end if

  return missingValue
Selesai
```

Gambar 2. *Pseudocode* Mencari Nilai yang Hilang

Pseudocode di atas mencari nilai yang hilang dalam struktur data yang terdiri dari kumpulan elemen (*array* 'arr'). Algoritma ini bekerja dengan mengiterasi melalui *array* dan memeriksa apakah nilai pada indeks 'ke-i' sama dengan 'i'. Jika tidak, itu berarti nilai yang hilang adalah 'i'. Jika tidak ada nilai yang hilang, nilai yang hilang adalah panjang *array* + 1.

```
Mulai
Mencari Duplikat(arr: Array of Integer)
  duplikat = Set()
  hasil = Set()

  for i from 0 to arr.length - 1
    if hasil.contains(arr[i])
      duplikat.add(arr[i])
    else
      hasil.add(arr[i])
    end if
  end for

  return duplikat
Selesai
```

Gambar 3. *Pseudocode* Mencari Duplikat Data

Pseudocode di atas mencari nilai duplikat dalam struktur data yang terdiri dari kumpulan elemen (*array* 'arr'). Algoritma ini menggunakan dua set 'hasil' untuk melacak nilai yang sudah ditemukan, dan 'duplikat' untuk menyimpan nilai duplikat yang ditemukan selama iterasi.

Setelah langkah-langkah tersebut, dataset diperiksa menyeluruh untuk memastikan dan mendeteksi nilai yang hilang dan duplikat. identifikasi nilai yang hilang dilakukan dalam setiap kolom, dan nilai yang hilang diperbaiki menggunakan metode imputasi

dengan nilai mean. selain itu, duplikat data juga dideteksi dan diatasi untuk memastikan integritas dan akurasi analisis.

Instances pada Data Set

Instances merujuk pada jumlah rekaman atau catatan yang terdapat dalam data set. Total Instance pada data set ini adalah sebanyak 72 data.

Penerapan Algoritma *Regresi Linear*

Dataset adalah kumpulan data yang digunakan untuk analisis atau penelitian tertentu. Dalam penelitian ini, data yang dikumpulkan mencakup penjualan mobil Toyota selama lima tahun, mulai dari 2018 hingga 2023. Sumber data ini adalah GAIKINDO, sebuah organisasi yang mengumpulkan dan menerbitkan data industri otomotif di Indonesia. Dengan menggunakan data yang valid dan terpercaya dari GAIKINDO, penelitian dapat memastikan keakuratan dan relevansi analisis yang dilakukan.

Langkah-langkah umum dalam penerapan *Regresi Linear* adalah:

1. Persiapan Data: Persiapkan data dengan membaginya menjadi data pelatihan (untuk membuat model) dan data uji (untuk mengevaluasi kinerja model).
2. Pilih Model: Pilih model *Regresi Linear* yang sesuai dengan data.
3. Latih Model: Gunakan data pelatihan untuk melatih model *Regresi Linear* dengan menyesuaikan koefisien garis regresi.
4. Evaluasi Model: Gunakan data uji untuk mengevaluasi kinerja model dengan metrik yang sesuai (misalnya, *RMSE*, *R-squared*, *MAE*).
5. Prediksi: Gunakan model yang dilatih untuk memprediksi nilai y untuk data baru atau data yang tidak terlihat sebelumnya.

Berikut adalah *Pseudocode* menggunakan algoritma *Regresi Linear*.

```

mulai
Function Regresi Linear
(data_pelatihan, data_uji):
#Langkah 1: Latih Model
model = Latih Model Regresi Linear
(data_pelatihan)

#Langkah 2: Evaluasi Model
evaluasi = Evaluasi Model Regresi Linear
(model, data_uji)

#Langkah 3: Prediksi
prediksi = Prediksi Regresi Linear
(model, data_baru)

return evaluasi, prediksi

Function Latih Model Regresi Linear
(data_pelatihan):
#Implementasi algoritma Regresi Linear
untuk melatih model
#Menggunakan metode kuadrat terkecil
(least squares method)
Model = Metode Kuadrat Terkecil
(data_pelatihan)

return model

Function Evaluasi Model Regresi Linear
(model, data_uji):
#Evaluasi kinerja model menggunakan
metrik yang sesuai
#Menghitung RMSE
evaluasi = Metrik Kinerja Model
(model, data_uji)

return evaluasi

Function Prediksi Regresi Linear
(model, data_baru):
# Gunakan model yang dilatih untuk
memprediksi nilai baru
prediksi = model.Prediksi(data_baru)

return prediksi
selesai

```

Gambar 4. *Pseudocode* Menggunakan Algoritma *Regresi Linear*

Metode *Least Square*

Metode *Least squares* adalah teknik statistik yang mencari garis atau kurva terbaik untuk data dengan meminimalkan jarak vertikal antara data dan garis atau kurva tersebut. Dalam peramalan penjualan, metode ini digunakan untuk melihat tren dari data penjualan masa lampau, membantu meramalkan penjualan di masa mendatang secara ringkas. Metode ini sesuai untuk mengidentifikasi hubungan linear antara variabel dependen dan satu atau lebih variabel independen. [22].

Data Latih dan Data Uji

Untuk membandingkan hasil prediksi yang diperoleh tingkat kesuksesan prediksi yang diperoleh tidak dapat dilihat. Oleh karena itu, untuk mengukur keberhasilan prediksi, akan disisihkan data 30% terakhir sebagai data uji dan 70% awal menjadi data latih. Jadi, penjualan mobil Toyota di Indonesia untuk beberapa bulan ke depan akan diprediksi berdasarkan data latih dan membandingkan hasilnya dengan nilai sebenarnya yang ada di data uji.

Penelitian ini menggunakan proporsi waktu sebagai cara untuk membagi data menggunakan train-test split. Dengan cara ini, memastikan bahwa model dilatih pada data historis yang lebih tua dan diuji pada data yang lebih baru, yang mencerminkan situasi dunia nyata di mana ingin melihat seberapa baik model, memprediksi masa depan berdasarkan informasi masa lalu. Jadi, dalam penelitian ini menggunakan 70% data untuk latihan dan 30% data untuk pengujian.

Berikut *Pseudocode* mencari data latih dan uji:

```

Mulai
1. Membaca Dataset
  a. Buka file dataset
  b. Baca data dari file dataset dan simpan dalam variabel 'dataset'
2. Membagi Dataset menjadi Data Latih dan Data Uji
  a. Tentukan proporsi data latih (misal: 0.7) dan data uji (misal: 0.3)
  b. Hitung jumlah total sampel dalam dataset
  c. Hitung jumlah sampel untuk data latih: total_sampel *
proporsi_data_latih
  d. Hitung jumlah sampel untuk data uji: total_sampel *
proporsi_data_uji
  e. Acak urutan dataset
  f. Ambil sejumlah sampel pertama sebagai data latih
  g. Ambil sisa sampel sebagai data uji
3. Mengembalikan Data Latih dan Data Uji
  a. Kembalikan data latih
  b. Kembalikan data uji
Selesai
  
```

Gambar 5. *Pseudocode* Data latih dan Data Uji

Penjelasan:

1. Membaca *Dataset*.

Buka file dataset Baca data dari file dataset.

2. Membagi *Dataset* menjadi Data Latih dan Data Uji

Tentukan proporsi data latih (70%) dan data uji (30%) Hitung jumlah total sampel dalam dataset. Hitung jumlah sampel untuk data latih total sampel * 0.7. Hitung jumlah sampel untuk data uji total sampel * 0.3. Ambil data 70% pertama sebagai data latih. Ambil data 30% terakhir sebagai data uji.

3. Mengembalikan Data Latih dan Data Uji

Kembalikan data latih Kembalikan data uji.

Dari data uji dan data latih yang diambil akan diprediksi 4 bulan kedepan.

Tabel 2. Data Uji (30%) Penjualan Mobil Toyota Tahun 2018-2023

No.	Bulan	Penjualan
1	2022-05	13297
2	2022-06	27290
3	2022-07	29326
4	2022-08	30844
5	2022-09	33449

6	2022-10	33740
7	2022-11	26462
8	2022-12	28128
9	2023-01	28970
10	2023-02	27336
11	2023-03	29471
12	2023-04	21518
13	2023-05	28178
14	2023-06	26092
15	2023-07	25908
16	2023-08	28104
17	2023-09	26600
18	2023-10	28003
19	2023-11	29125
20	2023-12	30825

Data latih terdiri dari 20 *Instances*. Setelah dilakukannya tahap *Preprocessing* telah dilakukan untuk memastikan integritas dan kelengkapan data sebelum melanjutkan analisis lebih lanjut.

Evaluasi

Dalam konteks penelitian, evaluasi kinerja adalah proses untuk menilai seberapa baik model yang dikembangkan dalam memprediksi hasil yang diinginkan. Metrik seperti MAE, MSE, dan RMSE digunakan untuk mengukur akurasi prediksi model dengan membandingkan hasil prediksi dengan data sebenarnya.

Mean Absolute Error (MAE)

MAE (Mean Absolute Error) adalah salah satu metrik evaluasi dalam machine learning yang digunakan untuk mengukur seberapa dekat prediksi model dengan nilai aktual. Metrik ini menghitung rata-rata dari nilai absolut selisih antara prediksi dan nilai sebenarnya, tanpa mempertimbangkan arah kesalahan (positif atau negatif). Rumus MAE adalah sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - F(X_i)|$$

Dimana:

- n adalah jumlah total observasi.
- Y_i adalah nilai aktual.
- $F(X_i)$ adalah nilai prediksi dari model untuk observasi ke-i.
- $Y_i - F(X_i)$ adalah selisih absolut antara nilai aktual dan nilai prediksi.
- Σ menunjukkan penjumlahan dari semua selisih absolut.
- $\frac{1}{n}$ adalah kebalikan dari jumlah total observasi, yang merupakan rata-rata dari selisih absolut.

MAE memberikan ukuran rata-rata dari kesalahan prediksi dalam skala yang sama dengan variabel yang diamati, dan semakin rendah nilai MAE, semakin baik modelnya dalam melakukan prediksi

Mean Squared Error (MSE)

MSE merupakan metrik yang menghitung rata-rata dari selisih kuadrat antara nilai prediksi dan nilai aktual. Rumus MSE adalah sebagai berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - F(X_i))^2$$

Dimana:

- n adalah jumlah total observasi.
- Y_i adalah nilai aktual.
- $F(X_i)$ adalah nilai prediksi dari model untuk observasi ke-i.
- $(Y_i - F(X_i))^2$ adalah selisih absolut antara nilai aktual dan nilai prediksi.
- Σ menunjukkan penjumlahan dari semua selisih absolut.
- $\frac{1}{n}$ adalah kebalikan dari jumlah total observasi, yang merupakan rata-rata dari selisih absolut.

MSE memberikan indikasi seberapa baik model memprediksi nilai aktual, dan semakin rendah nilai MSE, semakin baik pula performa model dalam melakukan prediksi.

Root Mean Squared Error (RMSE)

RMSE (Root Mean Squared Error) adalah salah satu metrik yang digunakan untuk mengukur seberapa baik model melakukan prediksi. Metrik ini dihasilkan dengan menghitung akar kuadrat dari nilai Mean Squared Error (MSE). Rumus RMSE adalah sebagai berikut:

$$RMSE = \sqrt{(MSE)}$$

Dimana:

MSE adalah Mean Squared Error.

$\sqrt{\quad}$ adalah fungsi akar kuadrat.

RMSE sering digunakan dalam evaluasi performa model prediksi karena memberikan gambaran yang intuitif, ukuran yang menunjukkan besarnya kesalahan dalam hasil prediksi. Nilai RMSE yang rendah atau mendekati angka 0 menunjukkan bahwa hasil peramalan sangat sesuai dengan data aktual. Meskipun RMSE biasanya berupa angka positif, semakin kecil nilai RMSE, semakin akurat prediksi yang dihasilkan. Berdasarkan penjelasan ini, model dengan RMSE dapat diimplementasikan dalam penelitian ini dan digunakan di masa mendatang untuk evaluasi dan perbaikan prediksi [23].

Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) adalah metrik yang digunakan dalam analisis statistik untuk menilai akurasi model prediksi atau peramalan. Metrik ini memberikan gambaran tentang seberapa akurat prediksi model jika dibandingkan dengan nilai aktual yang sebenarnya. [24].

$$MAPE = \sum_{t=1}^n \left[\frac{y_t - \bar{y}}{y_t} \right] \times 100\%$$

Dimana:

Σ menunjukkan penjumlahan dari semua selisih absolut.

n adalah jumlah total observasi.

t = 1 penjumlahan dimulai pada periode waktu dan berlanjut hingga periode waktu terakhir dalam kumpulan data.

$\frac{y_t - \bar{y}}{y_t}$ Persentase kesalahan antara nilai aktual dan nilai yang diprediksi pada periode atau data ke-i.

Nilai MAPE dapat diartikan ke dalam 4 kategori yaitu :

Tabel 3. Range Nilai *MAPE*

<i>Range MAPE</i>	Arti Nilai
< 10%	Kemampuan Model Peramalan Sangat Baik
10 - 20%	Kemampuan Model Peramalan Baik
20 - 50 %	Kemampuan Model Peramalan Layak
>50%	Kemampuan Model Peramalan Buruk

Interpretasi Hasil Evaluasi

Model dengan nilai MAE, MSE, RMSE, dan MAPE yang lebih kecil memiliki kinerja yang lebih baik. MAE mencerminkan rata-rata perbedaan antara nilai prediksi dan aktual, MSE mengukur rata-rata kuadrat perbedaan tersebut, RMSE menunjukkan rata-rata selisih dalam satuan yang sama dengan nilai aktual, dan MAPE merupakan metrik untuk menghitung rata-rata persentase kesalahan mutlak

Prosedur evaluasi model *Regresi Linear* dengan MAE, MSE, RMSE dan MAPE adalah sebagai berikut:

1. Pisahkan data menjadi data latihan dan data uji.
2. Latih model *Regresi Linear* dengan data latihan.
3. Prediksi nilai penjualan mobil Toyota dengan model yang telah dilatih.
4. Hitung nilai MAE, MSE, RMSE, dan MAPE antara nilai prediksi dan nilai aktual.

Dalam penelitian ini, data penjualan mobil Toyota akan dibagi menjadi data latihan dan data uji. Model *Regresi Linear* akan dilatih dengan data latihan dan kemudian digunakan untuk memprediksi nilai penjualan mobil Toyota untuk 4 bulan ke depan. Nilai MAE, MSE, RMSE, dan MAPE akan dihitung antara nilai prediksi dan nilai aktual.

```
Mulai
# Latih model Regresi Linear dengan
data training
model = regresi_linear(data_training)

# Prediksi nilai penjualan mobil Toyota
dengan model yang telah dilatih
prediksi = prediksi_penjualan
(model, data_testing)

# Hitung nilai MAE, MSE, dan RMSE antara
nilai prediksi dan nilai aktual
mae, mse, rmse = hitung_metrik
(prediksi, data_testing)

# Interpretasi hasil evaluasi
nilai_evaluasi = interpretasi_hasil
(mae, mse, rmse, mape)
Selesai
```

Gambar 6. *Pseudocode* Evaluasi dengan MAE, MSE, dan RMSE

Dengan demikian, hasil evaluasi model *Regresi Linear* dapat digunakan untuk menilai kinerja model dalam memprediksi penjualan mobil Toyota. Hasil evaluasi ini dapat digunakan sebagai dasar untuk meningkatkan kinerja model atau untuk menentukan apakah model dapat digunakan untuk memprediksi penjualan mobil Toyota di masa depan.

HASIL DAN PEMBAHASAN

Berdasarkan hasil dari pengolahan data penjualan mobil Toyota menggunakan metode *Regresi Linear* didapatkan hasilnya sebagai berikut.

Hasil Prediksi

Tabel 4. Hasil Prediksi dan Aktual Penjualan 2024

Tahun- Bulan	Prediksi Penjualan	Aktual Penjualan	Selisih Penjualan
2024-01	25356	24686	670
2024-02	25366	22142	3224
2024-03	25375	26142	-767
2024-04	25385	19577	5808

Hasil prediksi penjualan mobil Toyota untuk tahun 2024 menunjukkan angka yang cukup dekat dengan penjualan aktual pada bulan Januari 2024. Selisih ini menunjukkan bahwa model cukup akurat dalam memprediksi penjualan. Namun, pada bulan Februari, Maret, dan April 2024 menunjukkan adanya perbedaan signifikan antara prediksi dan realisasi, yang mungkin disebabkan oleh faktor-faktor musiman atau kejadian khusus yang tidak tertangkap oleh model. Hal ini mungkin terjadi karena adanya perubahan pasar yang tidak diantisipasi oleh model.

Hasil Evaluasi

Tabel 5. Hasil Nilai Error

No	Evaluasi	Nilai <i>Error</i>
1	MAE	2617.3161729166386
2	MSE	11291369.892151663
3	RMSE	3360.263366486571
4	MAPE	12.46934871611428

Dari evaluasi di atas, Model dengan Data Uji :

1. MAE memberikan rata-rata kesalahan absolut antara nilai aktual dan nilai prediksi. Dalam kasus ini, MAE adalah sekitar 2617 Ini berarti bahwa, rata-rata, prediksi model memiliki kesalahan sekitar 4946 unit penjualan.
2. MSE adalah rata-rata dari kesalahan kuadrat antara nilai aktual dan nilai prediksi. Dalam kasus ini, MSE adalah 11291369.89 adalah rata-rata dari kuadrat kesalahan

- antara prediksi dan nilai aktual untuk data masa depan.
3. RMSE sebesar 3360.26 adalah akar kuadrat dari MSE. Hasil penjualan 2,216.90 unit, RMSE yang lebih rendah untuk data masa depan menunjukkan bahwa model ini bekerja lebih baik dalam memprediksi data masa depan.
 4. Interpretasi MAPE sebesar 12.47%. menunjukkan bahwa rata-rata kesalahan relatif dari prediksi model adalah sekitar 12.47%. Ini berarti rata-rata perbedaan antara penjualan yang diprediksi dengan penjualan aktual adalah sekitar 12.47% dari nilai penjualan aktual. Semakin rendah nilai MAPE, semakin baik kualitas prediksi model.

SIMPULAN

Penerapan algoritma *Regresi Linear* dalam meramalkan penjualan mobil Toyota di industri otomotif Indonesia. Dengan menggunakan data penjualan dari tahun 2018 hingga 2023, model Regresi Linear berhasil memprediksi penjualan mobil Toyota untuk beberapa bulan ke depan. Hasil prediksi menunjukkan angka yang cukup dekat dengan penjualan aktual untuk bulan Januari, namun pada bulan Februari, Maret, dan April 2024 menunjukkan adanya perbedaan signifikan antara prediksi dan realisasi, yang mungkin disebabkan oleh faktor-faktor musiman atau kejadian khusus yang tidak tertangkap oleh model.

Evaluasi model menunjukkan bahwa Meskipun MAE menunjukkan kesalahan rata-rata yang relatif rendah, namun, MSE dan RMSE yang tinggi menandakan bahwa terdapat variasi yang besar antara prediksi dan nilai aktual, terutama pada bulan-bulan tertentu. Selain itu, nilai MAPE sebesar 12.47% mengategorikan model ini dalam kemampuan peramalan yang baik, sesuai dengan range nilai MAPE 10-20%. Ini menunjukkan bahwa rata-rata kesalahan relatif dari prediksi model adalah sekitar 12.47% dari nilai penjualan aktual.

DAFTAR PUSTAKA

- K. Yuliantari, N. Nurhidayati, and S. Sugiyah, "Brand Equity pada Mobil Toyota di Indonesia," *Moneter-Jurnal Akuntansi dan Keuangan*, vol. 9, no. 1, pp. 58–65, 2022.
- I. P. Tama, O. Novareza, D. Hardiningtyas, R. Yuniarti, and N. F. Nuzula, *Potensi Masa Depan Elektrifikasi Kendaraan Bermotor di Indonesia: Sebuah Analisis Strategis Rantai Pasok*. Universitas Brawijaya Press, 2023.

- T. Waluyo, A. Hermawan, and A. P. Wibowo, "Prediksi Penjualan Sepeda Motor Honda Menggunakan Jaringan Syaraf Tiruan," *Journal of Information System Management (JOISM)*, vol. 1, no. 1, pp. 31–35, Jul. 2019, doi: 10.24076/joism.2019v1i1.20.
- F. Riza, "Analisis dan Prediksi Data Penjualan Menggunakan Machine Learning dengan Pendekatan Ilmu Data," *Data Sciences Indonesia (DSI)*, vol. 1, no. 2, pp. 62–68, Jan. 2022, doi: 10.47709/dsi.v1i2.1308.
- D. Hatidja, F. G. Walelang, and C. E. J. C. Montolalu, "Penerapan Metode Autoregressive Integrated Moving Average (ARIMA) dalam Model Intervensi Fungsi Step untuk Memprediksi Penjualan Mobil di PT. Hasjrat Abadi Cabang Tendean Manado," *Indonesian Journal of Intelligence Data Science*, vol. 1, no. 1, 2023.
- N. Hajarisman, M. Herlina, and S. Stat, "Analisis Regresi dan Aplikasinya menggunakan SPSS." Program Studi Statistika, FMIPA Universitas Islam Bandung, 2022.
- S. Adiguno, Y. Syahra, and M. Yetri, "Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 1, no. 4, pp. 275–281, 2022.
- E. Junaedi, A. Siregar, and E. Nurlaelasari, "Implementasi C4. 5 Dan Algoritma K Nearest Neighbor Untuk Prediksi Kelayakan Pemberian Kredit Menggunakan RapidMiner Studio," *Scientific Student Journal for Information, Technology and Science*, vol. 3, no. 1, pp. 83–90, 2022.
- M. Lase, D. Saripurna, and V. W. Sari, "Estimasi Penjualan Ice Cream Walls Menggunakan Metode Regresi Linear Berganda," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 1, no. 5, pp. 625–634, 2022.
- I. L. L. Gaol, S. Sinurat, and E. R. Siagian, "Implementasi Data Mining Dengan Metode Regresi Linear Berganda Untuk Memprediksi Data Persediaan Buku Pada Pt. Yudhistira Ghalia Indonesia Area Sumatera Utara," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Nov. 2019, doi: 10.30865/komik.v3i1.1579.
- S. Dewi, H. A. Al Kautsar, and D. Y. Utami, "Prediksi Keberhasilan Pemasaran Layanan Jasa Perbankan Menggunakan Algoritma Logistic Regreesion," *Computer Science (CO-SCIENCE)*, vol. 3, no. 2, pp. 118–125, 2023.
- F. P. Dewanti, S. Setiyowati, and S. Harjanto, "Prediksi Persediaan Obat Untuk Proses Penjualan Menggunakan Metode Decision Tree Pada Apotek," *Jurnal Teknologi Informasi dan Komunikasi (TIKOMSiN)*, vol. 10, no. 1, May 2022, doi: 10.30646/tikomsin.v10i1.604.

- A. A.-F. Nur Wahyudin, A. Primajaya, and A. S. Y. Irawan, "Penerapan Algoritma Regresi Linear Berganda Pada Estimasi Penjualan Mobil Astra Isuzu," *Techno.Com*, vol. 19, no. 4, pp. 364–374, Nov. 2020, doi: 10.33633/tc.v19i4.3834.
- R. C. Rasyidi and I. G. B. B. Dharma, "Desain Eksperimen pada Sisi Mobil dengan Menggunakan Taguchi," *Seminar dan Konferensi Nasional IDEC*, 2019.
- A. Sudarismiati and M. T. Sari, "Analisis peramalan penjualan untuk menentukan rencana produksi pada UD Rifa'l," *Growth*, vol. 14, no. 2, pp. 17–30, 2019.
- C. Zai, "Implementasi Data Mining Sebagai Pengolahan Data," *Jurnal Portal Data*, vol. 2, no. 3, 2022.
- P. Susanti and K. Sussolaikah, "Penerapan metode regresi linear untuk memprediksi harga jual mobil bekas yaris dan jazz pada wilayah DKI Jakarta," *Network Engineering Research Operation*, vol. 7, no. 2, pp. 133–144, 2022.
- D. S. O. Panggabean, E. Bulolo, and N. Silalahi, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, p. 56, Feb. 2020, doi: 10.30865/jurikom.v7i1.1947.
- F. O. Suparno and S. Suwitho, "Pengaruh Harga, Lokasi dan Kualitas Produk Terhadap Keputusan Pembelian Perumahan Wisata Semanggi," *Jurnal Ilmu dan Riset Manajemen (JIRM)*, vol. 8, no. 5, 2019.
- T. Indarwati, T. Irawati, and E. Rimawati, "Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone," *Jurnal Teknologi Informasi dan Komunikasi (TIKoSIN)*, vol. 6, no. 2, pp. 2–7, 2019, doi: 10.30646/tikomsin.v6i2.369.
- Rudi Triatmono, "DATA PENJUALAN MOBIL TAHUN 2010 – 2023," <https://triatmono.info/data-penjualan-tahun-2012/data-penjualan-mobil-2017/>.
- F. R. Hariri, "Metode Least Square Untuk Prediksi Penjualan Sari Kedelai Rosi," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 7, no. 2, p. 731, Nov. 2016, doi: 10.24176/simet.v7i2.788.
- A. R. Kurniawan and N. A. Amri, "Estimasi Sumberdaya Emas Menggunakan Metode Ordinary Kriging Pada Pit X, Pt. Indo Muro Kencana, Kec. Tanah Siang, Kab. Murung Raya, Kalimantan Tengah," 2019.
- S. Fachid and A. Triayudi, "Perbandingan Algoritma Regresi Linier dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19," *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, p. 68, Jan. 2022, doi: 10.30865/mib.v6i1.3492.