

BAB 2

LANDASAN TEORI

2.1 *Model Supervised Learning*

Supervised learning adalah pendekatan pemodelan yang melatih model menggunakan data berlabel, di mana setiap observasi memiliki pasangan nilai *input* dan *output* yang telah diketahui, sehingga model dapat mempelajari pola hubungan antara keduanya untuk menghasilkan prediksi pada data baru.

Supervised learning dalam konteks peramalan (*forecasting*) merujuk pada pendekatan yang memanfaatkan data historis berlabel untuk membangun model yang mampu menghasilkan nilai perkiraan pada periode mendatang. Setiap observasi dalam data pelatihan memiliki nilai input berupa data deret waktu pada periode sebelumnya serta nilai *output* berupa nilai aktual pada periode berikutnya yang telah diketahui, sehingga model dapat mempelajari pola perubahan data dari waktu ke waktu secara eksplisit. Secara konseptual, akurasi peramalan model *supervised learning* sangat bergantung pada kemampuan data historis dalam merepresentasikan pola tren dan musiman yang terdapat dalam data aktual [21].

Dalam penelitian ini, *supervised learning* diimplementasikan melalui metode peramalan *Holt-Winters* yang dilatih pada data historis penjualan mobil bulanan berdasarkan kategori dari GAIKINDO periode 2013–2025, di mana setiap nilai penjualan pada periode tertentu digunakan untuk memperbarui komponen level, tren, dan musiman model secara bertahap melalui parameter pemulusan sehingga model mampu menyesuaikan diri terhadap perubahan pola data dari waktu ke waktu.

2.2 *Forecasting*

Forecasting berbasis data historis telah terbukti efektif diterapkan pada berbagai sektor industri di Indonesia, termasuk pada sektor energi untuk peramalan pembangkitan listrik nasional dengan tingkat akurasi yang tinggi [22]. Dalam konteks bisnis dan industri, *forecasting* sangat diperlukan khususnya pada industri otomotif Indonesia, karena fluktuasi yang terjadi pada penjualan mobil dapat mempengaruhi perencanaan produksi, pengelolaan stok, dan potensi pendapatan. Ketidakakuratan dalam peramalan dapat menyebabkan kelebihan stok (*overstock*) maupun kekurangan stok (*stockout*) [6]. Oleh karena itu, digunakan pendekatan

kuantitatif berbasis *time series* karena data yang digunakan merupakan data penjualan mobil GAIKINDO 2013 - 2025 berdasarkan kategori yang bersifat numerik historis dengan pola tren dan musiman.

Adapun beberapa metode *forecasting* lain yang sering digunakan dalam analisis data *time series* sebagai berikut.

1. ARIMA (*AutoRegressive Integrated Moving Average*): Cocok untuk pola linear sederhana, tapi memerlukan optimasi parameter yang kompleks.
2. LSTM (*Long Short-Term Memory*): *Deep Learning* untuk non-linier kompleks, membutuhkan sumber daya komputasi tinggi.

Studi komparatif yang menguji berbagai metode peramalan pada data deret waktu musiman menunjukkan bahwa metode berbasis *exponential smoothing* seperti *Holt-Winters* tetap kompetitif dibandingkan pendekatan *machine learning* maupun *deep learning* untuk data dengan pola musiman yang jelas, tanpa memerlukan kompleksitas komputasi yang tinggi [23]. Sebagai perbandingan, pendekatan *deep learning* seperti Bi-LSTM memang mampu menghasilkan akurasi yang tinggi pada data deret waktu finansial yang fluktuatif seperti data Forex, namun dengan arsitektur dan kompleksitas komputasi yang jauh lebih besar [24].

Salah satu contoh implementasi *forecasting* di sektor otomotif adalah untuk memprediksi arah penjualan mobil sebagai representasi dari permintaan pasar di masa mendatang [6]. Berdasarkan studi yang diimplementasikan pada data historis penjualan mobil GAIKINDO, pemodelan peramalan memberikan tingkat akurasi yang baik (*good forecasting*), dengan capaian nilai *Mean Absolute Percentage Error* (MAPE) sebesar 11,75% pada merek Toyota dan 15,02% pada merek Honda [6]. Dalam praktiknya, hasil prediksi yang akurat digunakan secara langsung oleh perusahaan sebagai pendukung keputusan (*decision support*), khususnya untuk merencanakan estimasi produksi, mengoptimalkan tenaga kerja yang dibutuhkan, serta mencegah penurunan pendapatan akibat kekurangan pasokan barang [6].

2.3 Time Series

Time series merupakan jenis data yang dicatat secara berurutan berdasarkan interval waktu tertentu. Dalam penelitian ini, data yang digunakan berupa data penjualan mobil bulanan berdasarkan kategori kendaraan yang bersumber dari laporan GAIKINDO periode 2013 - 2025. Karena data tersebut tersusun secara kronologis dari waktu ke waktu, maka pendekatan analisis *time series* digunakan

untuk memahami pola perubahan penjualan serta memodelkan kecenderungan penjualan pada periode mendatang.

Dalam analisis *time series*, pola data umumnya dibentuk oleh beberapa komponen utama yang menggambarkan karakteristik pergerakan data dari waktu ke waktu. Komponen tersebut meliputi tren, musiman, siklus, dan variasi acak. Identifikasi terhadap komponen-komponen ini penting karena metode peramalan yang digunakan harus mampu merepresentasikan pola yang terdapat dalam data historis.

1. *Trend* (T)

Menunjukkan kecenderungan arah pergerakan data dalam jangka panjang, baik meningkat, menurun, maupun relatif stabil. Pada data penjualan kendaraan, tren dapat mencerminkan perubahan permintaan pasar terhadap kategori kendaraan tertentu dari waktu ke waktu.

2. *Seasonality* (S)

Pola yang berulang secara periodik dalam interval waktu tertentu. Pada data penjualan mobil bulanan, pola musiman dapat muncul pada periode-periode tertentu seperti menjelang hari raya atau akhir tahun ketika permintaan kendaraan cenderung meningkat.

3. *Cyclical* (C)

Menggambarkan fluktuasi jangka panjang yang biasanya dipengaruhi oleh kondisi ekonomi makro atau dinamika industri. Perubahan kondisi ekonomi dapat menyebabkan naik turunnya permintaan kendaraan dalam periode yang lebih panjang.

4. *Irregular/Random* (I)

Variasi acak yang tidak dapat diprediksi secara sistematis dan biasanya muncul akibat faktor eksternal yang tidak terduga.

Dengan memahami komponen-komponen tersebut, analisis *time series* dapat digunakan untuk memodelkan pola historis data secara lebih sistematis. Dalam penelitian ini, karakteristik data penjualan mobil yang menunjukkan pola tren dan musiman menjadi dasar dalam pemilihan metode *Holt-Winters* sebagai model peramalan.

2.4 Exponential Smoothing

Exponential Smoothing adalah metode peramalan yang bekerja dengan cara merata-ratakan data historis secara berulang, namun memberikan bobot atau pengaruh yang lebih besar pada observasi terbaru [25]. Keunggulan utama dari metode ini adalah fleksibilitasnya yang tinggi, di mana metode ini mampu memproses data deret waktu yang berfluktuasi tanpa mengharuskan data tersebut berada dalam kondisi stasioner (stabil) [26]. Metode *Exponential Smoothing* disesuaikan menjadi tiga variasi berdasarkan karakteristik pola data yang diteliti.

1. Single Exponential Smoothing (SES)

Digunakan untuk meramalkan data yang berfluktuasi secara acak tanpa memiliki pola tren maupun musiman. Persamaan SES dirumuskan sebagai berikut.

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1} \quad (2.1)$$

$$\hat{y}_{t+h|t} = \ell_t \quad (2.2)$$

di mana ℓ_t merepresentasikan nilai level pada periode ke- t , y_t adalah nilai aktual pada periode ke- t , α adalah parameter pemulusan dengan rentang nilai $0 < \alpha < 1$, dan $\hat{y}_{t+h|t}$ adalah nilai ramalan untuk h periode ke depan dari periode t .

2. Double Exponential Smoothing (DES)

Digunakan apabila terdapat tren atau perubahan yang konsisten pada data, namun tidak memiliki pola musiman. Persamaan *Double Exponential Smoothing* atau *Holt's Linear Trend* dirumuskan sebagai berikut.

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (2.3)$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1} \quad (2.4)$$

$$\hat{y}_{t+h|t} = \ell_t + b_t h \quad (2.5)$$

di mana ℓ_t merepresentasikan nilai level pada periode ke- t , b_t menunjukkan nilai tren pada periode ke- t , α adalah parameter pemulusan level, β^* adalah parameter pemulusan tren, dan $\hat{y}_{t+h|t}$ adalah nilai ramalan untuk h periode ke depan dari periode t .

3. *Triple Exponential Smoothing (TES) atau Holt-Winters*

Digunakan ketika data memiliki pola tren sekaligus variasi musiman secara bersamaan. Uraian lengkap mengenai persamaan dan kedua variannya dijelaskan pada Subbab 2.5.

Pada penelitian ini, *Holt-Winters* dipilih karena pendekatan tersebut dinilai sesuai dengan karakteristik data historis penjualan mobil berdasarkan kategori dari GAIKINDO (2013–2025) yang secara aktual memiliki pola pergerakan tren dan musiman [6, 9]. Pemilihan ini didasarkan pada konsep teoretis metode *Holt-Winters* yang secara khusus memang direkomendasikan untuk memodelkan data deret waktu yang mengandung unsur tren serta fluktuasi musiman secara bersamaan [25, 26].

2.5 *Holt-Winters*

Holt-Winters adalah metode peramalan (*forecasting*) pada data *time series* yang memiliki pola tren dan musiman [27]. Metode ini dikembangkan berdasarkan konsep *exponential smoothing* oleh Holt (1957) dan disempurnakan oleh Winters (1960) [6]. Dengan menerapkan proses pemulusan secara berulang terhadap data historis, metode ini mampu menyesuaikan perubahan pola dari waktu ke waktu dan banyak digunakan dalam analisis ekonomi maupun prediksi penjualan.

Sebagai bentuk implementasi dari *Triple Exponential Smoothing* yang telah diuraikan sebelumnya, pemodelan *Holt-Winters* bekerja dengan memuluskan data historis melalui tiga parameter pemulusan. Ketiga parameter tersebut meliputi parameter α untuk level, β^* untuk tren, dan γ untuk musiman, dengan rentang nilai antara 0 hingga 1 [6]. Metode *Holt-Winters* didasarkan pada tiga komponen utama sebagai berikut [6].

1. *Level* (ℓ)

Merepresentasikan nilai dasar dari data pada periode tertentu yang telah disesuaikan dengan pengaruh tren dan musiman.

2. *Growth/Trend (b)*

Menunjukkan arah perubahan data dari waktu ke waktu, apakah mengalami kenaikan atau penurunan secara bertahap.

3. *Seasonality (s)*

Menggambarkan pola musiman yang berulang dalam periode tertentu, misalnya pola tahunan pada data penjualan bulanan.

Pada metode *Holt-Winters* terdapat dua metode yang berbeda, tergantung pada pemodelan musiman yang digunakan.

1. *Multiplicative Holt-Winters (MHW)*

Metode ini menunjukkan bahwa efek musiman meningkat seiring dengan meningkatnya level deret waktu.

$$\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (2.6)$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1} \quad (2.7)$$

$$s_t = \gamma \frac{y_t}{\ell_{t-1} + b_{t-1}} + (1 - \gamma)s_{t-m} \quad (2.8)$$

$$\hat{y}_{t+h|t} = (\ell_t + b_t h) s_{t-m+h_m^+} \quad (2.9)$$

Parameter m menyatakan panjang periode musiman. Komponen ℓ_t merepresentasikan level, b_t menunjukkan tren, dan s_t merupakan faktor musiman pada periode ke- t . Nilai $\hat{y}_{t+h|t}$ merupakan hasil peramalan dengan h sebagai periode ke depan.

2. *Additive Holt-Winters (AHW)*

Metode ini tidak bergantung pada level deret waktu saat ini dan komponen musimannya dapat dengan mudah ditambahkan atau dikurangkan dari hasil peramalan.

$$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (2.10)$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1} \quad (2.11)$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (2.12)$$

$$\hat{y}_{t+h|t} = \ell_t + b_t h + s_{t-m+h_m^+} \quad (2.13)$$

Perbedaan utama kedua model terletak pada bentuk operasi matematis yang digunakan dalam merepresentasikan efek musiman terhadap level deret waktu.

Pemilihan antara model *Additive* dan *Multiplicative* dilakukan berdasarkan perbandingan nilai *error* pada data aktual, di mana model dengan MAPE terkecil dipilih sebagai model terbaik untuk masing-masing deret waktu [28].

2.6 Seleksi Window Training

Window training dalam konteks peramalan deret waktu merujuk pada panjang data historis yang digunakan untuk melatih model peramalan pada satu periode tertentu. Dalam penelitian yang menggunakan data dengan karakteristik heterogen antar sub-kategori, panjang *window training* yang optimal tidak dapat diasumsikan seragam karena setiap deret waktu dapat memiliki pola tren dan musiman yang berbeda-beda.

Untuk menentukan panjang *window training* yang optimal, dapat digunakan pendekatan *rolling validation* atau yang dikenal juga sebagai *time series cross-validation*. Pendekatan ini merupakan salah satu teknik validasi yang direkomendasikan dalam pipeline peramalan deret waktu karena menjaga integritas kronologis data dan menghasilkan estimasi akurasi yang lebih representatif dibandingkan validasi statis [29]. Hyndman dan Athanasopoulos (2021) menyebut pendekatan ini sebagai *evaluation on a rolling forecasting origin* dan menyatakan bahwa cara yang baik untuk memilih model peramalan terbaik adalah dengan mencari model yang menghasilkan nilai RMSE terkecil yang dihitung menggunakan *time series cross-validation* [30].

2.7 Golden Section Search

Golden Section Search (GSS) adalah metode optimasi numerik untuk mencari nilai minimum suatu fungsi satu variabel secara iteratif dengan mempersempit interval pencarian menggunakan rasio emas $\phi = \frac{\sqrt{5}-1}{2} \approx 0,618$ [6]. Pada setiap iterasi, dua titik uji b dan c dievaluasi dalam interval $[a, d]$ menggunakan persamaan berikut.

$$b = \phi \cdot a + (1 - \phi) \cdot d \quad (2.14)$$

$$c = a + d - b \quad (2.15)$$

Apabila nilai fungsi objektif di titik b lebih kecil dari di titik c , maka interval diperbarui menjadi $[a, c]$. Sebaliknya, apabila nilai fungsi objektif di titik c lebih kecil, interval diperbarui menjadi $[b, d]$. Proses ini diulang hingga selisih $|d - a|$ lebih kecil dari nilai toleransi konvergensi ε yang telah ditentukan atau jumlah iterasi mencapai batas maksimum. Nilai optimal x^* kemudian dipilih dari kandidat terbaik di antara titik a , b , c , dan d yang menghasilkan nilai fungsi objektif terkecil.

2.8 Metrik Evaluasi Peramalan

Kinerja model peramalan dalam penelitian ini dievaluasi menggunakan tiga metrik kesalahan, yaitu *Mean Absolute Percentage Error* (MAPE), *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE).

Mean Absolute Percentage Error (MAPE) mengukur rata-rata persentase kesalahan absolut antara nilai aktual dan nilai prediksi. Metrik ini dipilih sebagai metrik utama karena hasilnya dalam bentuk persentase sehingga memudahkan perbandingan antar sub-kategori yang memiliki skala penjualan berbeda. MAPE dihitung hanya pada periode dengan nilai aktual tidak sama dengan nol untuk menghindari pembagian dengan nol. Keunggulan penggunaan MAPE adalah metrik ini dapat memberikan informasi mengenai tingkat ukuran kesalahan dalam peramalan, yang dapat mempermudah jika tingkat kesalahan peramalan terlalu rendah atau terlalu tinggi, sehingga hasil yang diperoleh lebih akurat [31]. MAPE dirumuskan sebagai berikut.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\% \quad (2.16)$$

Mean Absolute Error (MAE) mengukur rata-rata kesalahan absolut dalam satuan unit penjualan. Metrik ini memberikan gambaran yang intuitif mengenai seberapa besar rata-rata deviasi prediksi dari nilai aktual dalam satuan yang sama dengan data asli. MAE dirumuskan sebagai berikut.

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (2.17)$$

Root Mean Squared Error (RMSE) mengukur akar dari rata-rata kuadrat kesalahan. RMSE memberikan penalti kuadratik terhadap kesalahan besar sehingga lebih sensitif dalam mendeteksi deviasi ekstrem, sementara MAE lebih *robust* terhadap *outlier* karena menggunakan kesalahan absolut tanpa pembobotan [32]. *RMSE* dirumuskan sebagai berikut.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (2.18)$$

Pada ketiga persamaan di atas, Y_t merupakan nilai aktual penjualan pada periode t , $\hat{Y}_t$ merupakan nilai hasil prediksi model, dan n adalah jumlah periode yang dievaluasi. Semakin kecil nilai ketiga metrik tersebut, semakin baik kinerja model yang dihasilkan. Interpretasi nilai *MAPE* mengacu pada klasifikasi Lewis (1982) sebagaimana tercantum pada Tabel 2.1.

Tabel 2.1. MAPE Interpretation (Lewis, 1982)

MAPE	Interpretation
< 10	Highly Accurate forecasting
10 – 20	Good forecasting
20 – 50	Reasonable forecasting
> 50	Inaccurate forecasting

2.9 Penelitian Terdahulu

Tabel 2.2. Penelitian Terdahulu

No	Judul	Penulis	Algo/Metode	Hasil
1.	Parameter Optimization Of Holt-Winters Exponential Smoothing Using Golden Section Method for Predicting Indonesian Car Sales	Hani'ah et al. (2021)	Holt-Winters dengan optimasi Golden Section	Model Holt-Winters dengan optimasi Golden Section mencapai akurasi peramalan yang baik dengan nilai MAPE sebesar 11,75% untuk Toyota, 15,02% untuk Honda, dan 17,29% untuk Daihatsu.
2.	Forecast of the Evolution Trend of Total Vehicle Sales and Power Structure of China under Different Scenarios	Zhao et al. (2023)	Quadratic Exponential Smoothing, GM (1,1), PSO, Markov Model	Model peramalan gabungan yang dioptimasi dengan Particle Swarm mencapai nilai MAPE prediksi sebesar 0,705% dalam meramalkan total penjualan mobil.
3.	Forecasting of Automobile Sales Based on Support Vector Regression Optimized by the Grey Wolf Optimizer Algorithm	Qu et al. (2022)	Support Vector Regression (SVR), Grey Wolf Optimizer (GWO)	Model GWO-SVR mencapai performa prediksi penjualan mobil yang paling akurat dengan nilai MAPE sebesar 0,0239 pada merek Suteng dan 0,1419 pada merek Kaluola.
Lanjut ke halaman berikutnya				

Tabel 2.2 Lanjutan Penelitian Terdahulu

No	Judul	Penulis	Algo/Metode	Hasil
4.	Improving the Forecasting Performance of Taiwan Car Sales Movement Direction Using Online Sentiment Data and CNN-LSTM Model	Ou-Yang et al. (2022)	CNN-LSTM	Model deep learning CNN-LSTM yang dipadukan dengan data sentimen publik mencapai peningkatan akurasi peramalan arah penjualan mobil sebesar 27,78% dibandingkan model klasik.
5.	On Holt Winters Algorithm with Decomposition for Forecasting Financial Time Series with Complex Seasonal Patterns	Mona Nazih Ali Abdel Bary (2024)	Exponential Smoothing	Algoritma Holt-Winters terbukti sebagai model terbaik dalam memprediksi data deret waktu yang berfluktuasi dengan mencapai nilai kesalahan terendah yaitu MAPE sebesar 4, MAD sebesar 380, dan MSD sebesar 244.783.
6.	Assessing and Forecasting the Long-Term Impact of the Global Financial Crisis on New Car Sales in South Africa	Makoni & Chikobvu (2023)	SARIMA	Model SARIMA mencapai tingkat kesalahan yang sangat kecil dengan nilai MAPE sebesar 0,0003 dalam memprediksi tren jangka panjang pada fluktuasi penjualan mobil baru.
Lanjut ke halaman berikutnya				

Tabel 2.2 Lanjutan Penelitian Terdahulu

No	Judul	Penulis	Algo/Metode	Hasil
7.	Pengujian Model Multiplicative Holt-Winter's Exponential Smoothing dalam Peramalan Data Time-Series Terdampak Covid-19	Ayunda et al. (2022)	Multiplicative Holt-Winters	Model Multiplicative Holt-Winters mencapai akurasi prediksi yang sangat baik dengan nilai MAPE sebesar 9,98% pada data deret waktu yang memiliki pola tren musiman.
8.	Linear versus Exponential Forecasting Techniques: A Comparative Study of ARIMA and Holt-Winters Triple Exponential Smoothing Models for Indian Retail Gold Prices - Evidence from daily retail gold price data in India, 2014-2025	Rajib Bhattacharya (2025)	ARIMA, Holt-Winters Triple Exponential Smoothing	Model Holt-Winters mencapai nilai MAPE sebesar 4,87%, sedikit lebih akurat dan responsif dibandingkan metode linier ARIMA yang mencapai nilai MAPE sebesar 4,88%.
Lanjut ke halaman berikutnya				

Tabel 2.2 Lanjutan Penelitian Terdahulu

No	Judul	Penulis	Algo/Metode	Hasil
9.	Enhanced sales forecasting model using textual search data: Fusing dynamics with big data	A. Borah and O. Rutz (2024)	Machine Learning (Bayesian Elastic Net)	Model peramalan Machine Learning yang mengintegrasikan data pencarian daring kompetitor mencapai performa prediksi penjualan mobil yang jauh lebih presisi dibandingkan model dasar tanpa data pencarian.
10.	Machine Learning Fusion and Data Analytics Models for Demand Forecasting in the Automotive Industry: A Comparative Study	Kamal et al. (2023)	Random Forest, MLR, KNN, XGBoost, SVM	Model Support Vector Machine (SVM) mencapai tingkat akurasi tertinggi sebesar 71,42% dalam memprediksi permintaan merek otomotif, mengungguli Machine Learning lainnya seperti XGBoost yang hanya mencapai 42,85%.