

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian yang Relevan

Agar pembahasan penelitian ini tetap fokus dan terarah, ruang lingkup masalah dibatasi pada analisis tren harga minyak goreng menggunakan pendekatan deret waktu . Berbagai penelitian sebelumnya menunjukkan bahwa metode Linear Regression sering digunakan dalam analisis ekonomi dan komoditas karena kemampuannya dalam memodelkan hubungan antara variabel waktu dan harga secara sederhana serta mudah diinterpretasikan oleh pembuat kebijakan. Pendekatan tersebut memungkinkan peneliti untuk memperoleh gambaran yang jelas mengenai kecenderungan arah pergerakan harga berdasarkan data historis yang tersedia.

Untuk memastikan kebaruan dan posisi penelitian ini dalam peta keilmuan, berikut adalah ulasan terhadap penelitian terdahulu yang relevan dengan topik analisis tren harga komoditas pangan menggunakan metode *Regression Linear*:

Penelitian Pertama *"Assessing equivalence in raw accelerometer outputs across different brands using shaker table validation: a comparative analysis with filtering and linear regression techniques"* Penelitian tersebut mengkaji pemodelan harga minyak mentah Indonesia menggunakan pendekatan Regression linear dengan data bulanan periode 2018 - 2022, di mana komponen *Dated Brent* dan *Alpha* digunakan sebagai variabel pembentuk harga ICP. Pemodelan dilakukan dengan memanfaatkan variabel waktu, yaitu tahun dan bulan, untuk memprediksi masing-masing komponen harga yang selanjutnya digunakan dalam estimasi harga ICP. Hasil pengujian menunjukkan bahwa model *Regression linear* memiliki tingkat akurasi yang cukup baik, khususnya pada jenis minyak mentah SLC dengan nilai MAPE sebesar 9% yang tergolong sangat baik, sehingga dapat disimpulkan bahwa metode *Regression linear* mampu merepresentasikan kecenderungan pergerakan harga minyak secara kuantitatif dan menghasilkan estimasi yang relatif stabil untuk keperluan analisis dan peramalan (Coyle-Asbil et al., 2026)

Penelitian Kedua. ***"Estimation of withdrawal strength of different nail diameters in coconut (Cocos nucifera) lumber subjected to accelerated aging cycles using ultrasonic tests and multiple linear regression analysis"*** Penelitian tersebut membahas peramalan harga minyak mentah Indonesia menggunakan pendekatan Regression linear dengan memanfaatkan data bulanan periode 2018 - 2022. Data yang digunakan mencakup komponen Dated-Brent dan Alpha sebagai variabel pembentuk ICP, serta enam jenis minyak utama, yaitu SLC, Attaka, Duri, Belida, Banyu Urip, dan Senipah Condensate (SC). Pemodelan dilakukan dengan membangun persamaan Multiple Linear Regression untuk memprediksi nilai *Dated Brent* dan *Alpha* berdasarkan variabel waktu (tahun dan bulan), kemudian kedua hasil prediksi tersebut dijumlahkan untuk memperoleh estimasi harga ICP. Hasil evaluasi menunjukkan bahwa tingkat akurasi model bervariasi pada setiap jenis minyak, dengan minyak SLC memperoleh nilai MAPE sebesar 9% yang tergolong sangat baik. Temuan ini mengindikasikan bahwa metode *Regression linear* mampu menangkap pola tren harga minyak secara kuantitatif dan memberikan gambaran pergerakan harga yang relatif akurat, sehingga relevan dijadikan landasan metodologis dalam penelitian ini yang menganalisis tren harga minyak di Indonesia periode 2020–2025 menggunakan pendekatan Linear Regression (De Vera & Antonio, 2026).

Penelitian Ketiga ***"Comparative data-driven modeling of thermal energy storage using artificial neural networks and multiple linear regression"*** (Abdelhafez et al., 2026) Penelitian tersebut membandingkan berbagai pendekatan pemodelan, baik metode statistik konvensional maupun metode berbasis *machine learning*, untuk menilai kemampuan masing-masing model dalam menangkap pola tren dan fluktuasi harga minyak. Data yang digunakan berupa harga minyak mentah dari beberapa tahun terakhir yang dianalisis guna mengidentifikasi kecenderungan pergerakan harga dalam jangka pendek dan jangka panjang. Hasil penelitian menunjukkan bahwa model berbasis Regression masih menunjukkan kinerja yang kompetitif, khususnya pada data dengan pola linear. Temuan ini relevan dengan penelitian yang dilakukan karena sama-sama menggunakan data historis harga minyak sebagai variabel utama dan menitikberatkan pada analisis tren, terutama

melalui penerapan metode *Linear Regression* untuk menggambarkan arah pergerakan harga secara kuantitatif dan sistematis.

Penelitian Keempat ***"Penerapan Fuzzy Time Series dalam Peramalan Harga Minyak Serai pada CV Agam Jaya Atsiri"*** (Ella & Arifianto, 2020). Penelitian tersebut bertujuan memprediksi harga untuk satu bulan ke depan serta mengukur tingkat kesalahan prediksi menggunakan *Average Forecasting Error Rate (AFER)*. Proses analisis dilakukan secara bertahap, mulai dari penentuan himpunan semesta, pembentukan interval, fuzzifikasi, penyusunan *Fuzzy Logical Relationship Group (FLRG)*, hingga *defuzzifikasi* dan perhitungan nilai error. Hasil penelitian menunjukkan bahwa metode FTS mampu menghasilkan akurasi yang sangat baik dengan nilai AFER terendah sebesar 1,47%, sehingga dinilai efektif dalam memodelkan pola harga. Temuan ini relevan dengan penelitian yang dilakukan karena sama-sama memanfaatkan data runtun waktu dalam analisis tren harga komoditas, meskipun pendekatan yang digunakan dalam skripsi ini adalah *Regression linear* sebagai metode kuantitatif yang lebih sederhana dan mudah diinterpretasikan.

Penelitian Kelima ***"Prediksi Harga Minyak Mentah WTI Menggunakan Metode Fuzzy Time Series Markov Chain"***. Penelitian tersebut mengkaji peramalan harga minyak mentah jenis *West Texas Intermediate (WTI)* yang memiliki karakteristik fluktuatif dan berpengaruh terhadap kondisi ekonomi global. Data yang digunakan sebanyak 700 data runtun waktu pada periode 3 Januari 2021 hingga 6 April 2023, yang dibagi menjadi data pelatihan sebesar 75% (525 data) dan data pengujian sebesar 25% (175 data). Tahapan analisis meliputi pembentukan himpunan semesta, penentuan interval, proses fuzzifikasi, penyusunan *Fuzzy Logical Relationship (FLR)* dan *Fuzzy Logical Relationship Group (FLRG)*, pembentukan matriks probabilitas transisi Markov, hingga *defuzzifikasi* untuk memperoleh nilai prediksi akhir. Evaluasi model dilakukan menggunakan *Mean Absolute Percentage Error (MAPE)*, dengan hasil pengujian menunjukkan nilai MAPE sebesar 1,18% pada data testing dan tingkat akurasi lebih dari 98%. Hasil tersebut menandakan bahwa metode *Fuzzy Time Series Markov Chain* sangat baik

dan layak digunakan untuk prediksi jangka pendek, yang dalam penelitian tersebut dibatasi hingga 30 hari ke depan. Temuan ini menunjukkan bahwa pendekatan FTS *Markov Chain* efektif dalam memodelkan data harga minyak yang berfluktuasi, serta dapat menjadi alternatif yang andal dalam peramalan data ekonomi (Nurlela et al., 2023).

Penelitian Keenam ***”Peramalan Harga Minyak Mentah Dunia Menggunakan Metode Fuzzy Time Series Logika Singh”***. Penelitian tersebut mengkaji peramalan harga minyak mentah dunia jenis *West Texas Intermediate* (WTI) menggunakan metode *Fuzzy Time Series* (FTS) logika Singh, yang merupakan pengembangan dari metode FTS untuk meningkatkan akurasi pada tahap *defuzzifikasi*. Data yang digunakan berupa data sekunder harga minyak mentah WTI periode November 2014 hingga Juni 2022 sebanyak 92 observasi yang diperoleh dari Investing.com. Tahapan analisis meliputi analisis deskriptif, pembentukan himpunan semesta dan interval kelas dengan rumus *Sturges*, pendefinisian himpunan *fuzzy*, proses *fuzzifikasi* data historis, penyusunan *Fuzzy Logical Relationship* (FLR), hingga perhitungan nilai peramalan dan evaluasi akurasi menggunakan *Mean Absolute Percentage Error* (MAPE). Hasil penelitian menunjukkan bahwa metode FTS logika Singh menghasilkan nilai MAPE sebesar 0,30%, yang berada di bawah ambang 10% dan termasuk dalam kategori sangat baik. Temuan ini menunjukkan bahwa metode tersebut mampu mengikuti pola fluktuasi harga minyak mentah secara akurat dan layak digunakan untuk peramalan pada periode selanjutnya (Hasibuan et al., 2024).

Penelitian Ketujuh ***“Aplikasi Pendekatan Agglomerative Hierarchical Time Series Clustering untuk Peramalan Data Harga Minyak Goreng di Indonesia”***. Penelitian tersebut mengelompokkan provinsi di Indonesia berdasarkan pola runtun waktu harga minyak goreng menggunakan metode *Agglomerative Hierarchical Clustering* (AHC), kemudian melakukan peramalan pada tingkat kelompok. Data yang digunakan berupa data bulanan harga minyak goreng pasar tradisional di 34 provinsi periode Juli 2017 - Mei 2022. Hasil penelitian menunjukkan bahwa metode average linkage dengan ukuran kemiripan Euclidean Distance menghasilkan

pengelompokan terbaik, dan peramalan menggunakan Auto ARIMA memberikan tingkat akurasi yang cukup baik. Temuan ini menunjukkan bahwa pendekatan time series clustering efektif untuk analisis dan peramalan harga komoditas berbasis data runtun waktu (Zen et al., 2022).

Penelitian Kedelapan "***Indonesian Crude Oil Price (ICP) Prediction Using Multiple Linear Regression Algorithm***" (Suryani, Des; Fadhila, Mutia; Labellapansa, 2022). Penelitian tersebut membahas prediksi harga *Indonesian Crude Oil Price* (ICP) menggunakan metode *Multiple Linear Regression* pada data *time series* periode 2018–2022 yang bersumber dari Kementerian ESDM. Model dibangun dengan memprediksi komponen *Dated Brent* dan *Alpha*, yang kemudian dijumlahkan untuk memperoleh nilai ICP. Data dibagi menjadi 80% data pelatihan dan 20% data pengujian, serta melalui tahap *preprocessing* untuk menangani ketidaklengkapan data. Hasil evaluasi menggunakan MSE, RMSE, dan MAPE menunjukkan bahwa model memiliki tingkat akurasi sangat baik pada minyak SLC (MAPE 9%), cukup baik pada *Attaka*, *Belida*, dan SC, namun kurang optimal pada Duri dan Banyu. Temuan ini relevan dengan penelitian yang dilakukan karena sama-sama memanfaatkan data historis dan pendekatan kuantitatif dalam peramalan harga minyak.

Penelitian Kesembilan "***Model Prediksi Harga Bahan Bakar Minyak Terhadap USD Menggunakan Regression Linear***". Penelitian tersebut mengkaji pemodelan hubungan antara harga BBM jenis Pertalite dan nilai tukar USD menggunakan metode *Regression linear* dengan data *time series* tahunan selama periode 1990 - 2023. Data harga BBM diperoleh dari Pertamina, sedangkan data kurs USD bersumber dari Lokadata. Pemodelan dilakukan untuk menilai besarnya pengaruh kurs USD terhadap harga BBM maupun sebaliknya melalui persamaan *Regression linear* sederhana, dengan evaluasi akurasi menggunakan *Mean Absolute Percentage Error* (MAPE). Hasil penelitian menunjukkan tingkat kesalahan yang sangat rendah, dengan nilai MAPE sebesar 1,3836% untuk prediksi harga BBM dan 0,5493% untuk prediksi kurs USD, sehingga model dinilai memiliki akurasi yang sangat baik. Temuan ini relevan dengan penelitian yang dilakukan karena sama-

sama menggunakan pendekatan kuantitatif berbasis data *time series* dalam pemodelan variabel ekonomi yang saling berkaitan (Informasi et al., 2024).

Penelitian Kesepuluh ” ***Perbandingan Metode Regression Linear dan Exponential Smoothing dalam Memprediksi Harga Minyak Goreng Kemasan di Indonesia***” (Handayani et al., 2025). Penelitian tersebut membandingkan metode *Regression Linear* dan *Exponential Smoothing* (ETS) dalam memprediksi harga minyak goreng kemasan di 34 provinsi di Indonesia menggunakan data rata-rata tahunan periode 2020 - 2024 dari Badan Pangan Nasional. Evaluasi model dilakukan menggunakan MAE, MSE, dan RMSE, dengan hasil menunjukkan bahwa metode ETS dengan parameter α 0,6 memiliki tingkat akurasi yang lebih baik dibandingkan *Regression Linear*. Penelitian ini relevan dengan penelitian penulis karena sama-sama membahas peramalan harga komoditas berbasis data *time series* serta perbandingan metode untuk memperoleh model terbaik.

Penelitian Kesebelas ” ***Pendekatan Clustering Time Series pada Peramalan Harga Minyak Goreng***”. Penelitian ini mengkaji peramalan harga minyak goreng perdesaan di 32 provinsi Indonesia dengan membandingkan model ARIMA tanpa *clustering* dan ARIMA yang didahului proses *clustering time series*. Data yang digunakan berupa data harga bulanan periode 2009 - 2016 yang bersumber dari BPS, dengan pembagian data latih tahun 2009 - 2015 dan data uji tahun 2016. Clustering dilakukan menggunakan *Hierarchical Agglomerative Clustering* dengan *average linkage* dan ukuran jarak *Complexity Invariant Dissimilarity*. Evaluasi model dilakukan menggunakan MAPE serta *koefisien cophenetic* dan *silhouette*. Hasil penelitian menunjukkan bahwa ARIMA individual memberikan akurasi lebih tinggi, namun pendekatan *clustering* lebih efisien dari sisi waktu dan sumber daya. Penelitian ini relevan karena sama-sama membahas peramalan harga komoditas berbasis data *time series* dengan mempertimbangkan keseimbangan antara akurasi dan efisiensi model (Adinugroho, 2021).

Penelitian Keduabelas ” ***Pengembangan Pemodelan Harga Beras di Wilayah Indonesia Bagian Barat dengan Pendekatan Clustering Time Series***” (Wijaya & Ngatini, 2020). Penelitian tersebut mengembangkan model peramalan harga beras

dengan mengombinasikan ARIMA dan *clustering time series* menggunakan data harga beras bulanan periode 2010 - 2018 dari Badan Pusat Statistik untuk 18 provinsi di Indonesia bagian barat. *Clustering* dilakukan melalui *hierarchical clustering* dengan ukuran jarak *Dynamic Time Warping* dan penentuan jumlah kluster berdasarkan koefisien silhouette, yang menghasilkan dua kluster utama. Hasil evaluasi menggunakan MAPE menunjukkan bahwa pemodelan ARIMA pada level kluster lebih akurat dan efisien dibandingkan pemodelan pada level provinsi. Studi ini relevan karena sama-sama menerapkan pendekatan time series dalam peramalan harga komoditas dengan menekankan evaluasi akurasi sebagai dasar pemilihan model.

2.1.1 Matriks Penelitian

Tabel 2. 1 Matriks Penelitian

No	Nama Peneliti	Judul Penelitian	Sumber Data	Klasifikasi	Arsitektur	Kesimpulan
1	Coyle-Asbil et al. (2026)	<i>Assessing equivalence in raw accelerometer outputs across different brands using shaker table validation: a comparative analysis with filtering and linear regression techniques</i>	<i>Data bulanan harga minyak mentah Indonesia (ICP) periode 2018–2022</i>	Regression Linear Time Series	Regression linear berbasis variabel waktu untuk memprediksi komponen pembentuk ICP	Akurasi cukup baik, terutama pada SLC (MAPE 9%).
2	De Vera & Antoni	<i>Estimation of withdrawal strength</i>	<i>Data ICP 2018–2022</i>	Multiple Linear Regression berbasis	<i>Prediksi Dated-Brent dan Alpha</i>	Model cukup akurat dan mampu

No	Nama Peneliti	Judul Penelitian	Sumber Data	Klasifikasi	Arsitektur	Kesimpulan
	o (2026)	<i>of different nail diameters in coconut lumber using ultrasonic tests and multiple linear regression analysis</i>		Time Series	<i>dengan MLR</i>	menangkap tren harga.
3	Abdelhafez et al. (2026)	<i>Comparative data-driven modeling of thermal energy storage using artificial neural networks and multiple linear regression</i>	<i>Data historis harga minyak beberapa pa tahun terakhir</i>	ANN dan Multiple Linear Regression	Perbandingan performa model statistik dan machine learning	Regression tetap kompetitif pada pola linear.
4	Ella & Arifianto (2020)	<i>Penerapan Fuzzy Time Series dalam Peramalan Harga Minyak Serai pada CV Agam Jaya Atsir</i>	Data harga minyak serai	<i>Fuzzy Time Series</i>	<i>Tahapan FTS hingga evaluasi AFER</i>	AFER 1,47% (sangat baik).
5	Nurlela et al. (2023)	<i>Prediksi Harga Minyak Mentah WTI Menggunakan</i>	Data WTI 2021–2023	FTS Markov Chain	Fuzzifikasi, FLR, FLRG, matriks Markov	MAPE 1,18%, akurat untuk jangka pendek.

No	Nama Peneliti	Judul Penelitian	Sumber Data	Klasifikasi	Arsitektur	Kesimpulan
		<i>kan Metode Fuzzy Time Series Markov Chai</i>				
6	Hasibuan et al. (2024)	<i>Peramalan Harga Minyak Mentah Dunia Menggunakan Metode Fuzzy Time Series Logika Singh</i>	Data WTI 2014–2022	FTS Logika Singh	Interval Sturges, fuzzifikasi, FLR	MAPE 0,30% (sangat baik).
7	Zen et al. (2022)	Aplikasi Pendekatan Agglomerative Hierarchical Time Series Clustering untuk Peramalan Data Harga Minyak Goreng di Indonesia	Data harga minyak goreng 2017–2022	Time Series Clustering & ARIMA	AHC dan Auto-ARIMA	Clustering efektif dan peramalan cukup akurat.
8	Suryani, Fadhila, & Labellapansa (2022)	Indonesian Crude Oil Price (ICP) Prediction Using Multiple Linear	Data ICP 2018–2022	Multiple Linear Regression	Prediksi komponen ICP	Akurasi sangat baik pada SLC (MAPE 9%).

No	Nama Peneliti	Judul Penelitian	Sumber Data	Klasifikasi	Arsitektur	Kesimpulan
		Regression Algorithm				
9	Informasi et al. (2024)	Model Prediksi Harga Bahan Bakar Minyak Terhadap USD Menggunakan Regressio n Linear	Data BBM dan USD 1990–2023	Regression Linear Sederhana	Pemodelan hubungan dua variabel ekonomi	MAPE sangat rendah (<2%), model sangat akurat.
10	Handayani et al. (2025)	Perbandin gan Metode Regressio n Linear dan Exponenti al Smoothing dalam Mempredi ksi Harga Minyak Goreng Kemasan di Indonesia	Data harga 2020–2024 (Badan Pangan Nasion al)	Regression Linear & ETS	Perbandinga n dua metode (MAE, MSE, RMSE)	ETS ($\alpha=0,6$) lebih akurat dibanding Regression linear.
11	Adinugroho (2021)	Pendekata n Clustering Time Series pada Peramalan Harga Minyak Goreng	Data harga bulanan minyak goreng perdesa an 32 provinsi periode 2009–	Time Series Clustering dan ARIMA	Hierarchical Agglomerati ve Clustering (average linkage, Complexity-Invariant Dissimilarit y), kemudian pemodelan	ARIMA individual memberikan akurasi lebih tinggi, namun pendekatan clustering lebih efisien dari sisi waktu dan sumber daya.

No	Nama Peneliti	Judul Penelitian	Sumber Data	Klasifikasi	Arsitektur	Kesimpulan
			2016 (BPS)		ARIMA; evaluasi dengan MAPE, cophenetic, dan silhouette	
12	Wijaya & Ngatini (2020)	Pengembangan Pemodelan Harga Beras di Wilayah Indonesia Bagian Barat dengan Pendekatan Clustering Time Series	Data harga beras bulanan 2010–2018 (Badan Pusat Statistik) untuk 18 provinsi di Indonesia bagian barat	Time Series Clustering dan ARIMA	Hierarchical clustering dengan jarak Dynamic Time Warping (DTW), penentuan jumlah kluster menggunakan koefisien silhouette (2 kluster), kemudian pemodelan ARIMA pada level kluster dan dibandingkan dengan ARIMA per provinsi; evaluasi menggunakan MAPE	ARIMA berbasis kluster menghasilkan akurasi yang lebih baik dan lebih efisien dibandingkan pemodelan pada level provinsi.

2.2 Landasan Teori

Landasan teori membahas tentang beberapa teori yang mendasari analisis permasalahan dan sesuai dengan topik yang dibahas, berikut adalah teori pada penelitian ini:

2.2.1 Harga Minyak Goreng

Minyak goreng merupakan salah satu komoditas pangan strategis yang menjadi kebutuhan pokok masyarakat Indonesia. Sebagai komoditas yang dikonsumsi secara luas oleh seluruh lapisan masyarakat, stabilitas harga minyak goreng memiliki dampak signifikan terhadap tingkat inflasi dan kesejahteraan rumah tangga (Zen et al., 2022). Tingginya konsumsi minyak goreng setiap tahunnya diiringi dengan meningkatnya permintaan dari pasar, sehingga menyebabkan harga minyak goreng cenderung meningkat tiap tahunnya (Zen et al., 2022).

2.2.2 Outlier pada Data Harga

Outlier adalah observasi yang menyimpang jauh dari pola umum data dan sering muncul dalam dataset harga yang volatile, sehingga dapat memengaruhi estimasi parameter dan kualitas prediksi model.” Mendukung laporan ini, (Harga & Rawit, 2026) menyatakan bahwa deteksi *outlier* penting dalam analisis harga karena *outlier* berdampak signifikan pada dinamika model volatil.

Dalam studi komoditas, (Harga & Rawit, 2026) menemukan bahwa *outlier* dalam data harga cabai harus dideteksi dan diperbaiki, karena nilai ekstrem secara substansial memengaruhi akurasi model prediksi harga. Penelitian (Trydini et al., 2023) juga menunjukkan bahwa deteksi *outlier* pada data harga saham memastikan estimasi model lebih robust dan prediksi lebih akurat.

2.2.3 Linear Regression

Regression Linear merupakan pendekatan statistik yang efektif untuk mengidentifikasi pola tren jangka panjang (Starbuck, 2023). Dalam konteks analisis tren harga komoditas, model ini memperlakukan waktu sebagai variabel independen dan harga sebagai variabel dependen, sehingga mampu menangkap arah pergerakan harga secara kuantitatif tanpa memerlukan kompleksitas model lanjutan seperti ARIMA atau *exponential smoothing*.

Bentuk umum model *Regression Linear* untuk analisis tren dinyatakan sebagai:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \quad (2.1)$$

di mana:

y_t = harga minyak goreng harian (Rp) pada hari ke - t

t = indeks waktu harian (1,2,3, ..., n)

β_0 = intercept (estimasi harga pada 1 Januari 2020)

β_1 = koefisien slope (laju perubahan harga per hari dalam $\frac{Rp}{hari}$)

ε_t = residual (penyimpangan acak dari tren)

Parameter β_0 dan β_1 diestimasi menggunakan metode *Ordinary Least Squares* (OLS) yang meminimalkan jumlah kuadrat residual:

$$\min \beta_0, \beta_1 \sum_{t=1}^n (y_t - \beta_0 - \beta_1 t)^2 \quad (2.2)$$

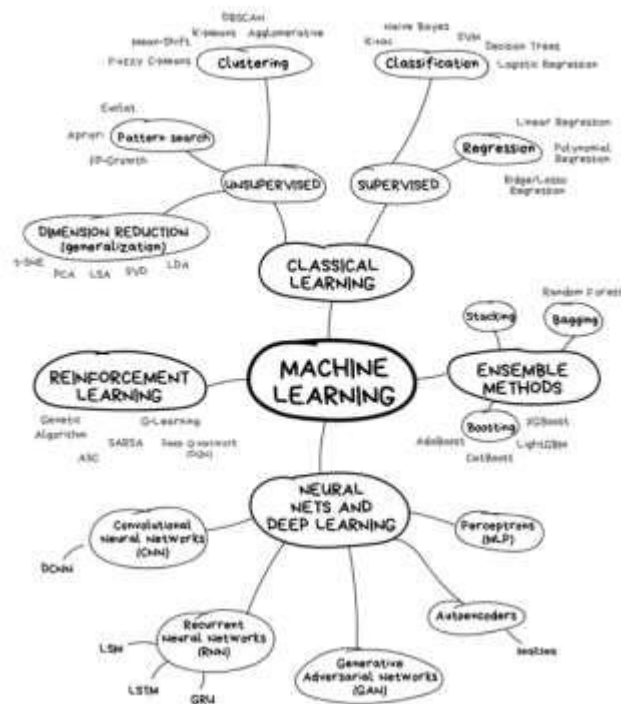
Pendekatan ini dipilih karena kesederhanaan interpretasinya dan kemampuannya merepresentasikan tren jangka panjang meskipun data harian mengandung fluktuasi jangka pendek (Starbuck, 2023). Koefisien β_1 secara langsung mengindikasikan:

- a. Nilai positif: tren kenaikan harga
- b. Nilai negatif: tren penurunan harga
- c. Nilai mendekati nol: stabilitas harga

2.2.4 Machine Learning

Machine Learning (ML) merupakan sub-bidang ilmu komputer yang memungkinkan sistem untuk belajar dari data tanpa diprogram secara eksplisit (Pengantar et al., 2024). Tujuan utamanya adalah mengembangkan algoritma yang mampu mengenali pola dalam data historis dan menghasilkan prediksi atau keputusan berdasarkan pola tersebut. Proses pembelajaran pada ML

melibatkan tiga komponen kunci: (1) representasi data, (2) fungsi evaluasi, dan (3) algoritma optimisasi (Pengantar et al., 2024). Dalam konteks analisis tren harga komoditas, Machine Learning berperan sebagai pendekatan kuantitatif untuk mengidentifikasi pola historis dan memproyeksikan kecenderungan pergerakan harga ke depan.



Gambar 2. 1 Machine Learning (Pengantar et al., 2024)

2.2.5 Asumsi Regression Linear

Dalam Regression linear berbasis *least squares*, keabsahan inferensi statistik seperti interval kepercayaan dan nilai p sangat bergantung pada terpenuhinya asumsi terhadap residual model. Barker dan Shaw menekankan bahwa agar *least squares regression* menghasilkan interval kepercayaan dan nilai p yang valid, residual harus memenuhi tiga asumsi utama: (1) residual saling independen atau tidak berkorelasi, (2) residual berdistribusi normal, dan (3) varians residual konstan (*homoskedastisitas*) (Watson et al., 2023). Jika salah satu asumsi ini dilanggar, estimasi koefisien dapat menjadi kurang akurat dan hasil uji statistik berpotensi menyesatkan, sehingga kesimpulan model

menjadi kurang dapat diandalkan. Dalam konteks data runtun waktu seperti harga minyak periode 2020 - 2025, asumsi independensi sering kali sulit dipenuhi karena observasi yang berurutan cenderung menimbulkan autokorelasi, sementara dinamika volatilitas harga dapat menyebabkan varians residual berubah-ubah (*heteroskedastisitas*). Oleh sebab itu, sebelum menarik kesimpulan terkait tren harga dari hasil *Regression*, perlu dilakukan analisis residual secara menyeluruh misalnya melalui inspeksi plot residual dan uji diagnostik yang relevan serta penerapan penyesuaian yang tepat apabila ditemukan pelanggaran asumsi, agar interpretasi tren tetap kuat dan defensif secara statistik.

2.2.6 *Preprocessing Data*

Preprocessing data merupakan tahapan penting dalam pemodelan prediksi karena mempersiapkan dataset agar terhindar dari *noise*, *missing value*, dan *outlier* yang dapat mengganggu performa model regresi linier serta memperbaiki kualitas input agar estimasi dan prediksi lebih akurat (Koukaras & Tjortjis, 2025).

Penerapan teknik preprocessing seperti pembersihan data, penanganan nilai yang hilang, dan normalisasi terbukti memperbaiki performa model regresi linier dalam prediksi harga saham pada data finansial (Linier et al., 2023).

2.2.7 *Pembagian Data Latih dan Data Uji*

Dalam penelitian machine learning, pembagian dataset menjadi data latih dan data uji merupakan prosedur metodologis yang penting untuk memastikan evaluasi model dilakukan secara objektif dan tidak bias. Data latih digunakan untuk mengestimasi parameter model dan menangkap pola dari data historis, sedangkan data uji berfungsi sebagai data independen untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum pernah :digunakan sebelumnya. Tanpa pemisahan ini, model berpotensi mengalami overfitting, yaitu kondisi ketika model terlalu menyesuaikan diri terhadap data latih sehingga performanya menurun saat diterapkan pada data baru. Studi empiris menunjukkan bahwa variasi rasio pembagian data, seperti

70:30 atau 80:20, dapat memengaruhi nilai evaluasi model secara signifikan (Bichri et al., 2024). Rasio data latih yang terlalu kecil dapat menyebabkan model kurang optimal dalam menangkap pola, sedangkan data uji yang terlalu sedikit dapat menghasilkan estimasi kinerja yang kurang stabil dan kurang representatif. Selain itu, penelitian di bidang pemrosesan citra medis juga menunjukkan bahwa pemilihan rasio pembagian dataset memengaruhi kemampuan generalisasi model serta tingkat akurasi klasifikasi (Sivakumar et al., 2024). Oleh karena itu, pemilihan rasio pembagian data harus mempertimbangkan ukuran sampel, kompleksitas model, serta tujuan analisis agar hasil evaluasi dapat dipertanggungjawabkan secara ilmiah dan memiliki daya generalisasi yang baik.

2.2.8 Evaluasi Kinerja Model

Evaluasi kinerja model bertujuan untuk mengukur seberapa dekat hasil prediksi dengan nilai aktual, sehingga kualitas pemodelan dapat dinilai secara kuantitatif dan objektif. (Jierula et al., 2021) menjelaskan bahwa metrik yang umum digunakan dalam evaluasi model Regression atau peramalan meliputi *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE) (Jierula et al., 2021) MSE dan RMSE memberikan penalti lebih besar terhadap kesalahan yang ekstrem karena menggunakan kuadrat selisih, sehingga keduanya sensitif terhadap deviasi besar MAE menghitung rata-rata kesalahan absolut sehingga lebih mudah diinterpretasikan secara langsung sementara MAPE menyajikan kesalahan dalam bentuk persentase, sehingga memudahkan perbandingan akurasi antarperiode maupun antarseri data. Oleh karena itu, dalam analisis tren harga minyak periode 2020 - 2025, penggunaan kombinasi metrik seperti MAE atau RMSE bersama MAPE memberikan gambaran evaluasi yang lebih komprehensif, karena model dinilai tidak hanya berdasarkan besarnya selisih prediksi dalam satuan harga, tetapi juga berdasarkan tingkat kesalahan relatif terhadap nilai aktualnya.

2.2.9 Koefisien Determinasi (R^2) dan Uji Signifikansi Parameter

Koefisien determinasi (R^2) digunakan untuk menunjukkan sejauh mana model Regression mampu menjelaskan variasi pada variabel dependen melalui variabel independen yang dimasukkan ke dalam model. Roustaei dkk. menjelaskan bahwa R^2 merepresentasikan proporsi variasi pengamatan yang dapat dijelaskan oleh model Regression, sehingga ukuran ini kerap dijadikan indikator *goodness-of-fit* pada *Regression linear* (Roustaei, 2024). Namun, kecocokan model saja belum cukup signifikansi statistik koefisien Regression juga perlu diuji untuk memastikan bahwa hubungan linear yang diestimasi tidak semata-mata muncul karena kebetulan sampel. Pengujian ini umumnya dilakukan melalui nilai p pada koefisien *Regression*, yang secara praktik setara dengan uji t untuk masing-masing koefisien (Roustaei, 2024) Dalam penelitian tren harga minyak periode 2020 - 2025, R^2 digunakan untuk menilai seberapa besar variasi harga dapat dijelaskan oleh variabel waktu (indeks periode), sedangkan uji signifikansi pada koefisien *slope* bertujuan untuk memastikan apakah arah tren baik kenaikan maupun penurunan benar-benar bermakna secara statistik. Jika koefisien *slope* terbukti signifikan, maka perubahan rata-rata harga per periode yang ditangkap oleh garis Regression dapat diperlakukan sebagai tren yang lebih kuat secara ilmiah, bukan sekadar fluktuasi acak semata.

2.2.10 Visualisasi Data

Visualisasi data merupakan proses mengubah data numerik atau teks menjadi bentuk grafik yang lebih mudah dipahami, sehingga pola, tren, hubungan antar variabel, dan anomali dapat diidentifikasi secara cepat dan akurat (**SITASI 1**). Dalam konteks *machine learning*, visualisasi menjadi alat penting selama tahapan eksplorasi data (*exploratory data analysis*), pra-pemrosesan, hingga interpretasi hasil model, karena membantu peneliti dalam memilih fitur, mengetahui sebaran data, serta mengidentifikasi outlier yang dapat memengaruhi performa model (Saxena & M.P, 2023). Penelitian sistematis terbaru juga menunjukkan bahwa visualisasi tidak hanya meningkatkan pemahaman terhadap data yang kompleks tetapi juga memainkan peran penting dalam *AI-assisted decision-making*, di mana berbagai bentuk

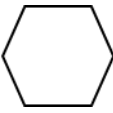
representasi visual seperti grafik, dashboard, dan plot interaktif memperkuat kemampuan pengguna dalam mengekstraksi wawasan dan membuat keputusan berbasis data (Neri et al., 2025). Kejelasan dalam visualisasi data memainkan peran ganda yaitu sebagai alat evaluasi internal bagi peneliti dan sebagai media komunikasi hasil kepada pemangku kepentingan non-teknis, sehingga meningkatkan efektivitas pengambilan keputusan dalam berbagai domain aplikasi (*data science*, kesehatan, dan kebijakan publik) (Neri et al., 2025; Saxena & M.P, 2023).

2.2.11 Flowchart

Flowchart adalah diagram alir yang dimanfaatkan untuk menggambarkan urutan proses atau logika dalam sebuah system secara visual. Dalam konteks penelitian berbasis kecerdasan buatan, *Flowchart* berperan dalam mengilustrasikan tahapan kerja sistem, mulai dari penumpulan data, pra-pemrosesan, pelatihan model evaluasi, hingga interpretasi hasil. Setiap Langkah direpresentasikan dengan symbol standarseperti oval (mulai/akhir), persegi panjang (proses), belah ketupat (keputusan), dan panah (arah aliran data) (Sutanti et al., 2020).

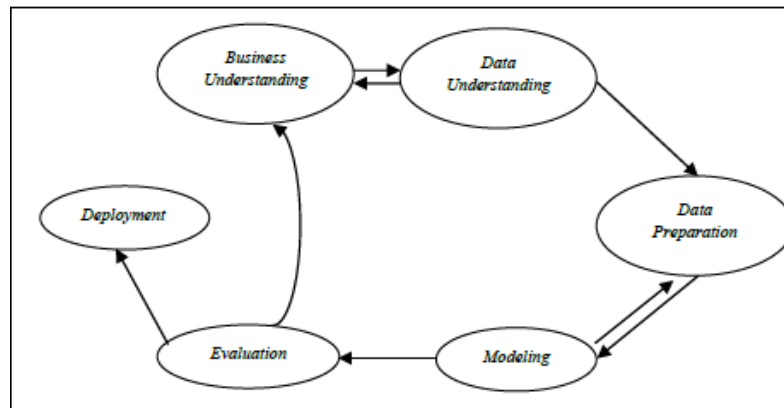
Tabel 2. 2 Flowchart

No	Simbol	Nama	Keterangan
1		Flow Direction	Digunakan untuk menghubungkan simbol yang satu dengan simbol yang lainnya
2		Terminator	Untuk permulaan (start) dan akhir (stop) dari suatu kegiatan
3		Processing	Simbol yang menunjukkan bahwa komputer melakukan pengolahan
4		Decision	Simbol proses yang dipilih berdasarkan kondisi saat ini

No	Simbol	Nama	Keterangan
5		Connector	Simbol penyambungan proses dalam lembar / halaman yang sama
6		Preparation	Simbol untuk menyiapkan penyimpanan yang akan diproses di dalam penyimpanan
7		Input-Output	Simbol yang menunjukkan proses input dan output terlepas dari jenis peralatan
8		Predefine Proses	Simbol yang menunjukkan pelaksanaan komponen atau prosedur tertentu
9		Display	Simbol yang menunjukkan peralatan output yang digunakan, seperti printer, plotter, layar, dan sebagainya
10		Punch Card	Simbol yang menunjukkan bahwa output ditulis ke kartu atau input berasal dari kartu

2.2.12 CRISP-DM

CRISP-DM (*Cross Industry Standard Process for Data Mining*) merupakan model proses standar dan independen industri yang banyak digunakan dalam proyek data mining dan data science. Model ini menyediakan alur kerja sistematis agar proses analisis data berjalan terstruktur dan terarah (Schröer et al., 2021).



Gambar 2. 2 Phase Of CRISP-DM (Dhewayani et al., 2022)

Gambar 2.2 menggambarkan enam tahapan inti dari CRISP-DM yang saling terkait. Penjelasan setiap tahapannya adalah sebagai berikut:

A. Business Understanding

Pada langkah pertama *business understanding* proses untuk menetapkan tujuan bisnis, memahami situasi dan kondisi yang ada saat penelitian, serta mendefinisikan tujuan penelitian dalam konteks masalah yang diatasi melalui data mining (Dhewayani et al., 2022).

B. Data Understanding

Tahap persiapan yang melibatkan pemeriksaan data yang digunakan, pengumpulan data awal, serta identifikasi kualitas data. Pada tahap ini, data akan menjalani proses deskripsi untuk setiap fiturnya (Dhewayani et al., 2022).

C. Data Preparation

Tahap persiapan yang melibatkan pemeriksaan data yang digunakan, pengumpulan data awal, serta identifikasi kualitas data. Pada tahap ini, data akan menjalani proses deskripsi untuk setiap fiturnya (Dhewayani et al., 2022).

D. Modelling

Tahap pelaksanaan algoritma yang digunakan untuk mencari, mengidentifikasi, serta menghasilkan pola pada data penelitian (Dhewayani et al., 2022).

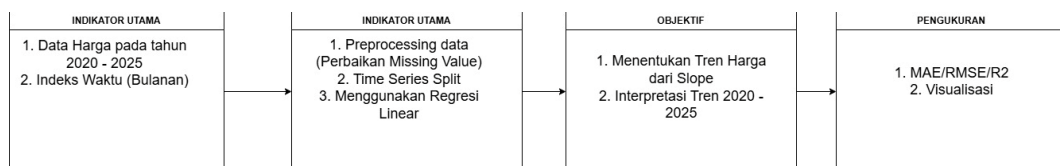
E. *Evaluation*

Proses untuk mengukur hasil dari model yang telah diimplementasikan pada tahap modeling sebelumnya. Hasil evaluasi ini menggambarkan proses data mining yang telah dilaksanakan serta menentukan model terbaik yang sesuai untuk digunakan (Dhewayani et al., 2022).

F. *Deployment*

Proses penyusunan laporan atau artikel jurnal berdasarkan hasil penelitian yang telah diperoleh (Dhewayani et al., 2022).

2.3 Kerangka Pemikiran



Gambar 2. 3 Kerangka Pemikiran

Gambar 2.3 menggambarkan kerangka pemikiran penelitian yang akan dilakukan, dengan penjelasan sebagai berikut:

1. Indikator

Indikator dalam penelitian ini adalah data harga harian minyak goreng di Provinsi DKI Jakarta pada periode Januari 2020 hingga Desember 2025 yang diperoleh dari Pusat Informasi Harga Pangan Strategis Nasional (PIHPS) Bank Indonesia melalui laman <https://www.bi.go.id/hargapangan/TabelHarga/PasarTradisionalKomoditas>. Data harian ini mencerminkan harga minyak goreng di pasar tradisional dan modern di wilayah DKI Jakarta, yang selanjutnya menjadi dasar analisis tren.

2. Metode Usulan

Penelitian ini menggunakan pendekatan pemodelan tren melalui *Regression Linear*, dengan waktu (dalam bentuk indeks periode harian) sebagai variabel independen dan harga minyak goreng sebagai variabel dependen. Sebelum model dibangun, data terlebih dahulu melalui tahap preprocessing yang meliputi:

- a. Pengecekan kelengkapan data harian
- b. Penanganan *missing values* pada hari libur atau akhir pekan melalui teknik *forward-fill* atau interpolasi linier
- c. Identifikasi dan penanganan *outlier* menggunakan analisis Z-score ($|Z| > 3$)

Setelah proses pembersihan data selesai, dataset dibagi menggunakan pendekatan *time-based split*, yaitu pemisahan berdasarkan urutan kronologis menjadi data latih (periode awal) dan data uji (periode akhir). Strategi ini dipilih agar proses evaluasi lebih mencerminkan kondisi nyata dalam analisis tren historis, di mana model dilatih menggunakan data historis harian dan kemudian dievaluasi terhadap periode harian berikutnya untuk mengukur kemampuan representasi tren jangka panjang.

3. Objektif

Tujuan penelitian ini adalah mengidentifikasi kecenderungan (tren) pergerakan harga minyak goreng di DKI Jakarta pada periode 2020 - 2025 secara kuantitatif melalui parameter dalam model Regression Linear. Fokus utamanya terletak pada:

- a. Koefisien *slope* (kemiringan), yang menunjukkan arah sekaligus laju perubahan harga per periode harian apakah cenderung meningkat (nilai positif), menurun (nilai negatif), atau stabil (mendekati nol).
- b. Nilai *intercept* yang merepresentasikan estimasi harga pada titik awal skala waktu (Januari 2020).

Penelitian ini tidak ditujukan untuk peramalan jangka panjang, melainkan untuk menganalisis dan merepresentasikan pola tren historis harga minyak goreng di DKI Jakarta secara objektif dan terukur sebagai dasar pertimbangan kebijakan stabilitas harga pangan di tingkat provinsi.

4. Pengukuran

Evaluasi kinerja model dilakukan dengan mengukur tingkat kesalahan prediksi pada data uji menggunakan tiga metrik yang telah ditetapkan dalam rumusan masalah:

- a. *Mean Absolute Error* (MAE): Mengukur rata-rata selisih absolut antara nilai prediksi dan aktual.
- b. *Root Mean Square Error* (RMSE): Mengukur akar rata-rata kuadrat selisih; lebih sensitif terhadap *outlier*.
- c. *Koefisien Determinasi* (R^2): Mengukur proporsi variasi harga yang dapat dijelaskan oleh variabel waktu (rentang 0 - 1; semakin mendekati 1 semakin baik).

Selain metrik kuantitatif, penelitian ini juga memanfaatkan visualisasi berupa:

- a. *Line chart* data harian dengan garis tren hasil *Regression Linear* untuk menggambarkan arah kecenderungan jangka panjang.