Jurnal Kecerdasan Buatan dan Teknologi Informasi
Vol. 5, No. 3, September 2026, pp. 782–790
e-ISSN: 2964-2922, p-ISSN: 2963-6191

Open access

Original Research

Fashion Product Sales Prediction Using Random Forest Regression Algorithm Based on Historical Data

Lusiana Situmorang [a], Mukhlis [b*], Muhammad Yasir [c]

[a,b,c] Informatics, Faculty of Computer Science, Universitas Bhayangkara Jakarta Raya, Bekasi, Indonesia

*Corresponding Author: mukhlis@dsn.ubharajaya.ac.id

ABSTRACT

Tulip Store at Metropolitan Mall Bekasi, managed by PT Star Asia Brothers, relies on historical sales data for evaluation and target planning. However, sales volume fluctuations, ranging from 385 units in October 2025 to 1,122 units in March 2026, make heuristic planning unreliable. This study builds a daily sales volume prediction model using Random Forest Regression based on daily recapitulation data from October 2025 to March 2026. Following the CRISP-DM framework, data preparation involved cleaning, daily aggregation, and feature engineering to construct 11 input features categorized into temporal indicators, historical lag variables, and rolling statistics. The model was trained using an 80:20 train-test split with hyperparameters $n\_estimators = 200$, $max\_depth = None$, and $random\_state = 42$. Evaluation on 36 test instances yielded a Mean Absolute Error (MAE) of 1.7649 and a Root Mean Square Error (RMSE) of 2.9740, with 91.67% of predictions exhibiting errors within ± 5 units. Feature importance analysis revealed that sequential month (*Bulan_ke*, 0.385) and previous-day sales (*Lag_1*, 0.294) were the most influential variables, indicating that monthly seasonal trends and immediate sales momentum drive daily demand. The proposed model provides robust operational decision support for daily sales evaluation, target setting, and stock allocation.

How to Cite:
L. Situmorang, Mukhlis, and M. Yasir, "Fashion Product Sales Prediction Using Random Forest Regression Algorithm Based on Historical Data," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 5, no. 3, pp. 782–790, 2026.
Doi: https://doi.org/10.69916/jkbti.v5i3.583

KEYWORDS

Sales Prediction
Random Forest Regression
Machine Learning
CRISP-DM
Feature Engineering

ARTICLE INFO

Received: July 13, 2026
Accepted: September 7, 2026
Available Online: September 19, 2026

1 INTRODUCTION

The fashion industry is one of the retail sectors with highly dynamic demand characteristics. Changes in consumer preferences, seasonal factors, trend developments, and social media influence cause fashion product demand patterns to shift rapidly, making the forecasting process more challenging. The relatively short product lifecycle requires businesses to accurately estimate market needs to avoid both excess and shortage of inventory [1]. The high volume of sales data generated from fashion retail activities encourages the use of data-driven approaches as more effective solutions in supporting demand forecasting and business decision-making [2].

PT Star Asia Brothers (SAB) is a retail company and womens undergarment supplier headquartered in West Jakarta. The company manages four undergarment brands: Tulip, Showy, Glena, and Christine [3]. One of the Tulip brand outlets operated by PT SAB is located at Metropolitan Mall (MM) Bekasi. All sales activities are conducted using a Point of Sales (POS) system, so every transaction is digitally recorded and compiled into daily sales reports. These reports contain information such as transaction date, product article code, basic indicator, color, product status, quantity sold (Qty), and product size.

Based on interviews with the Tulip Store Supervisor, historical sales data is used as a basis for sales evaluation, target setting, and operational strategy determination. However, the company faces difficulties in estimating sales conditions for upcoming periods, as sales volume fluctuations are influenced by various factors including seasonal sales, promotional programs, bulk purchases, and changes in consumer demand. Analysis of daily sales recapitulation data for the period October 2025 to March 2026 shows that total sales changed from 385 units in October 2025 to 1,122 units in March 2026, with February 2026 sales of only 391 units jumping to 1,122 units the following month.

Various previous studies have demonstrated that the Random Forest Regression algorithm can produce sales predictions with good accuracy across different domains. Falah et al. [4] showed that Random Forest performed better than linear regression in retail sales prediction. Hidayat et al. [5] confirmed that Random Forest Regression can improve sales prediction accuracy on supermarket data. Efendi and Zyen [6] proved that Random Forest can be effectively used not only for predicting sales but also for supporting product inventory analysis. Kurniawan et al. [7] successfully predicted daily Indomaret sales revenue with MAE = 9,587.48 and $R^2 = 0.9998$. Verdiyanto et al. [8] developed a POS-based application using Random Forest Regression for daily beverage sales prediction, achieving $R^2 = 0.9091$ with MAPE = 2.85%. Della et al. [9] specifically added Lag Feature and Day of Week features to Random Forest Regression for daily canteen sales prediction, achieving MAE = Rp23,150.84, RMSE = Rp28,905.32, and $R^2 = 0.9015$.

Despite these achievements, most previous studies were conducted in general retail contexts such as supermarkets, minimarkets, food businesses, or manufacturing products. Research applying prediction models to daily sales recapitulation data for fashion products, particularly undergarments, remains relatively limited. Furthermore, most studies focus primarily on model performance measurement, while the utilization of prediction results as supporting information in the sales evaluation process and target setting at the operational store level is rarely discussed.

This study builds a daily sales volume prediction model using the Random Forest Regression algorithm based on daily sales recapitulation data from Tulip Store at Metropolitan Mall Bekasi for the period October 2025 to March 2026. The novelty of this research lies in: (1) integrating daily temporal features and historical sales features as input variables in a Random Forest Regression model specifically designed for womens undergarment product sales data; (2) evaluating the model using MAE and RMSE metrics to produce more accurate and measurable daily sales volume predictions; and (3) positioning the prediction results as decision support for operational-level sales evaluation and target setting.

2 RESEARCH METHODS

2.1 Research Framework

This study employs the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology as the research framework for building the daily sales volume prediction model. CRISP-DM consists of six stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [4]. In this study, only the first five stages are applied, as the research focuses on model development and evaluation without deploying the model into an operational system [11–15].

2.2 Dataset

The dataset used in this study consists of daily sales recapitulation data from Tulip Store at Metropolitan Mall Bekasi for the period October 2025 to March 2026, comprising 1,403 transactions stored in six Microsoft Excel files. Each file represents one month of sales records. Only the Tanggal (Date) and Qty (Quantity) attributes are used in the modeling process. The Date attribute serves as the basis for temporal feature formation, while Qty is used to construct the target variable Total_Qty and historical features.

Table 1. Dataset Characteristics

No	Characteristic	Description
1	Data Source	Tulip Store MM Bekasi
2	Data Period	October 2025 – March 2026
3	Number of Files	6 Microsoft Excel files
4	Total Transactions	1,403 transactions
5	Attributes Used	Tanggal (Date) and Qty
6	Target Variable	Total_Qty (daily)

2.3 Data Preparation

The data preparation phase processes raw transaction data into an optimal structure for predictive modeling through feature engineering. In this step, new features are extracted and constructed from historical records to help the model identify temporal patterns, historical dependencies, and short-term trends.

Feature engineering produced 11 input variables organized into three categories: (a) temporal features extracted from the date column—*Hari* (day of month), *Hari_Minggu* (day of week, 0=Monday to 6=Sunday), *Minggu_ke* (week number within month), *Is_Weekend* (binary indicator), and *Bulan_ke* (sequential month number, 1=October through 6=March); (b) lag features—*Lag_1*, *Lag_2*, and *Lag_7* representing *Total_Qty* values from 1, 2, and 7 days prior, respectively; (c) rolling statistics—*Rolling_3* and *Rolling_7* representing 3-day and 7-day moving averages of *Total_Qty*; and (d) *Lag_Transaksi* representing the number of transactions from the previous day. The engineered features are categorized into three primary types, as summarized in Table 2.

Table 2. Feature Engineering Summary

No	Feature	Type	Description
1	Hari	Temporal	Day of month (1–31)
2	Hari_Minggu	Temporal	Day of week (0–6)
3	Minggu_ke	Temporal	Week number (1–5)
4	Is_Weekend	Temporal	Weekend indicator (0/1)
5	Bulan_ke	Temporal	Month sequence (1–6)
6	Lag_Transaksi	Historical	Transaction count $(t-1)$
7	Lag_1	Historical	Total_Qty $(t-1)$
8	Lag_2	Historical	Total_Qty $(t-2)$
9	Lag_7	Historical	Total_Qty $(t-7)$
10	Rolling_3	Rolling	3-day moving average
11	Rolling_7	Rolling	7-day moving average

Missing values generated by lag and rolling feature creation on initial rows were handled by dropping rows with NaN values, as these rows lacked sufficient historical data.

2.4 Modeling

The dataset was split into training and testing sets using the hold-out method with an 80:20 ratio via the `train_test_split()` function with `random_state = 42` to ensure reproducibility [10]. The Random Forest Regression model was built with the following parameters: `n_estimators = 200` (number of decision trees), `max_depth = None` (unlimited tree depth), `max_features = None` (all features considered at each split), and `random_state = 42`.

The prediction value from Random Forest Regression is obtained by averaging the outputs of all individual decision trees, as expressed in Equation (1):

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b(x) \tag{1}$$

where $\hat{y}$ is the final prediction, B is the number of decision trees, $T_b(x)$ is the prediction of the b-th tree, and x is the input data.

2.5 Evaluation Metrics

Model performance was evaluated using two metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). MAE measures the average absolute difference between actual and predicted values, while RMSE measures the square root of the average squared prediction errors, giving larger penalty to larger errors. Additionally, feature importance analysis, scatter plots, and error distribution histograms were employed for comprehensive model assessment [16–22].

3 RESULTS AND DISCUSSION

3.1 Business Understanding

The business understanding phase was conducted through direct observations and interviews with the Supervisor of Tulip Store at Metropolitan Mall Bekasi. Operating as a specialized retail outlet under PT Star Asia Brothers, all sales transactions are recorded via a Point of Sales (POS) system and automatically compiled into daily sales reports. These reports serve as the primary resource for monitoring sales achievement, evaluating store performance, and establishing sales targets.

Currently, sales targets are set using heuristic approaches, primarily relying on sales volumes from previous periods or same-period historical records from the prior year. Weekly evaluation meetings are routinely conducted to assess target fulfillment, evaluate store traffic, and formulate operational strategies, such as product replenishment or inter-store transfers. However, store management encounters significant operational challenges in accurately forecasting short-term sales volume. Daily sales volumes experience high instability, such as a sharp increase from 391 units in February 2026 to 1,122 units in March 2026, driven by promotional campaigns, seasonal buying behaviors, bulk purchases, and sudden shifts in consumer demand. A purely reactive approach to sales tracking hinders proactive stock allocation and target planning, reinforcing the necessity for a data-driven predictive model.

3.2 Data Understanding and Preparation Results

The initial dataset consisted of 1,403 raw transaction records spanning six months (October 2025 through March 2026) across six monthly Excel files. Monthly transaction density varied from 194 transactions in October 2025 to 313 transactions in March 2026. Data audit and cleaning procedures confirmed zero missing values, zero duplicate entries, and zero invalid transaction quantities.

To establish a uniform time series, transaction-level data was aggregated by date, reducing the dataset from 1,403 individual transactions into 177 distinct daily observation records. Feature engineering was subsequently performed on the aggregated daily data, creating 11 predictor features categorized into temporal, historical (lag), and rolling statistics.

Due to the time-shifted calculations of lag and moving average features (3-day and 7-day rolling windows), initial rows containing incomplete historical windows generated non-applicable entries. These rows were dropped to preserve data integrity, yielding a clean dataset for model execution.

3.3 Model Evaluation Results

The Random Forest Regression model was trained on an 80:20 split of the prepared dataset and evaluated against hold-out testing data, consisting of 36 daily test instances. The hyperparameter configuration was set to $n\_estimators = 200$, $max\_depth = None$, $max\_features = None$, and $random\_state = 42$. The model's predictive performance was quantified using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), as detailed in Table 3.

Table 3. Model Evaluation Results

Metric	Value
MAE (Mean Absolute Error)	1.7649
RMSE (Root Mean Square Error)	2.9740

The evaluation yielded an MAE of 1.7649, indicating that the model's daily volume predictions deviate from actual sales quantities by an average of approximately 1.76 units per day. The RMSE value of 2.9740, while slightly higher due to the quadratic penalization of larger variances, demonstrates low overall error variance across the evaluation window.

An in-depth analysis of individual error residuals across the 36 test data points reveals high prediction consistency:

- 33 out of 36 predictions (91.67%) exhibited absolute prediction errors within ± 5 units.
- 3 out of 36 predictions (8.33%) exceeded the ± 5 unit error threshold, primarily corresponding to unexpected sales spikes driven by bulk purchases.

The alignment between predicted and actual values confirms that the ensemble tree structure effectively captures underlying daily sales dynamics without overfitting to random fluctuations.

3.4 Feature Importance Analysis

Feature importance analysis revealed the relative contribution of each input variable to the prediction model. The results showed that *Bulan_ke* (sequential month number) and *Lag_1* (previous days sales) were the two most influential features. The dominance of *Bulan_ke* indicates that the month-level trend plays a significant role in determining daily sales volume, likely reflecting seasonal sales patterns observed in the data where sales volume varied substantially across months. The high importance of *Lag_1* confirms strong temporal dependency in daily sales data, consistent with findings by Della et al. [9] who also identified *Lag_1* as the most dominant predictor in daily sales prediction.

The remaining temporal features (*Hari_Minggu*, *Minggu_ke*, *Is_Weekend*, *Hari*) also contributed to the model but with lower importance scores, suggesting that within-month temporal patterns provide supplementary

predictive information. Rolling statistics features (*Rolling_3*, *Rolling_7*) captured short-term and medium-term sales trends, while *Lag_2*, *Lag_7*, and *Lag_Transaksi* provided additional historical context.
To quantify the relative contribution of each input variable to the Random Forest Regression model, a feature importance analysis was conducted based on the mean decrease in impurity. The 11 engineered features were evaluated to identify the key drivers influencing daily sales volume predictions. The relative importance scores for all features are illustrated in Figure 1.

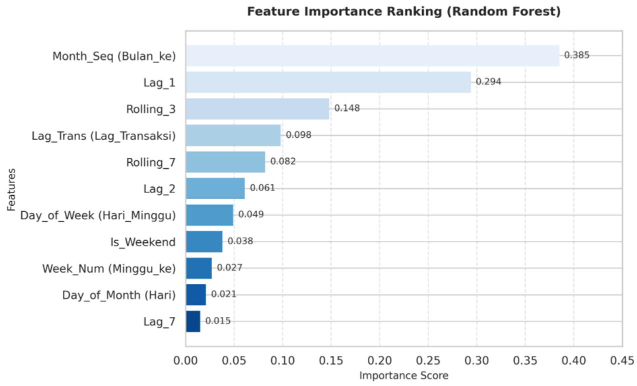

Figure 1. Feature Importance Ranking of the Random Forest Regression Model

As presented in Figure 1, the feature importance analysis reveals that Month_Seq (*Bulan_ke*) and *Lag_1* are the two most dominant predictors in the model:

- **Macro-Seasonal Impact (Month_Seq / *Bulan_ke*):** Achieving the highest importance score of 0.385, Month_Seq serves as the primary driver for daily sales volume. This demonstrates that broader monthly seasonal trends heavily influence overall product demand patterns.
- **Immediate Temporal Dependency (*Lag_1*):** With an importance score of 0.294, *Lag_1* ranks as the second most influential feature. This highlights strong short-term auto-correlation, indicating that sales performance on the preceding day $(t-1)$ provides a crucial baseline for predicting sales on day t.
- **Mid-Term Dynamics & Rolling Features:** *Rolling_3* (0.078), *Lag_Trans* (0.052), and *Rolling_7* (0.048) provide supporting historical context, helping the model smooth short-term fluctuations and adjust to multi-day transaction momentum.
- **Calendar & Micro-Temporal Signals:** Micro-temporal variables such as *Day_of_Week* (0.041), *Is_Weekend* (0.035), *Week_Num* (0.034), and *Day_of_Month* (0.033) contribute supplementary predictive value, primarily refining predictions for specific day-of-week behaviors.

3.5 Comparison with Previous Studies

To evaluate the predictive capability of the proposed model within a broader research context, the performance metrics obtained in this study were benchmarked against previous implementations of Random Forest algorithms across different retail domains. Table 4 presents a comparative summary of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values reported in related literature alongside the findings of this study.

Table 4. Comparison with Related Studies

Study	Method	MAE	RMSE	Domain
Falah et al. [4]	RF	4.121	5.246	Perfume retail
Hidayat et al. [5]	RFR	0.931	–	Supermarket
Della et al. [9]	RFR	23,150*	28,905*	Canteen (*Rp)
This study	**RFR**	**1.765**	**2.974**	**Fashion retail**

As summarized in Table 4, direct quantitative comparison across studies requires careful consideration of variations in target variables, operational domains, and data scaling:

- **Domain-Specific Precision:** Compared to general retail applications such as perfume sales prediction by Falah et al. [4] (MAE = 4.1210, RMSE = 5.2460), the proposed Random Forest Regression model achieved lower error metrics (MAE = 1.765, RMSE = 2.974). This indicates that integrating lag variables with temporal indicators effectively reduces prediction variance in the highly dynamic fashion retail environment.
- **Contextual Evaluation:** While Hidayat et al. [5] reported a lower MAE of 0.9310 in a supermarket setting, their study evaluated daily unit movements with different demand stability characteristics. Furthermore, studies like Della et al. [9] predicted sales in monetary values (IDR) rather than product units, highlighting

the necessity of assessing performance within domain-specific contexts.
- **Model Viability:** Overall, achieving an MAE of 1.765 units and an RMSE of 2.974 units confirms that the developed model delivers competitive accuracy, providing a reliable quantitative basis for daily demand forecasting in fashion retail outlets.

3.6 Practical Implications and Limitations

a Operational Decision Support

The predictive model provides direct operational value to store management at Tulip Store MM Bekasi across three main operational areas:

- **Sales Target Setting:** Replaces intuitive target setting with data-backed daily baselines, reducing unrealistic sales goals during low-demand periods.
- **Inventory & Stock Planning:** Provides store supervisors with advance estimates of daily unit movement, optimizing stock replenishment schedules from the central warehouse and preventing stockouts or over-inventory.
- **Staffing & Operational Allocation:** Allows store managers to anticipate high-volume sales days (indicated by *Is_Weekend* combined with high *Rolling_3* trends) and adjust shift schedules accordingly.

b Research Limitations

Despite its predictive accuracy, several constraints should be considered:

- **Time-Horizon Constraints:** The model was trained on a six-month dataset (October 2025 – March 2026). While sufficient for capturing short-term patterns, it lacks full annual coverage to model major yearly holiday cycles (e.g., Eid al-Fitr or Year-End sales campaigns).
- **Exogenous Factors:** External variables such as store discount rates, mall foot traffic, competitor promotional events, and macro-economic factors were not included in the dataset.
- **Algorithm Scope:** Evaluation was restricted to Random Forest Regression without cross-benchmarking against gradient boosting frameworks (e.g., XGBoost, LightGBM, or CatBoost).

3.7 Discussion

The empirical findings of this study demonstrate that the Random Forest Regression algorithm, combined with hybrid feature engineering, is highly effective for forecasting daily sales volumes in fashion retail, specifically within the women's undergarment segment. The model achieved a Mean Absolute Error (MAE) of 1.7649 and a Root Mean Square Error (RMSE) of 2.9740, with 91.67% of test predictions falling within an error margin of ± 5 units. These results highlight the model's capacity to learn non-linear relationships and temporal dependencies in short-term sales data without overfitting to random daily noise.

The feature importance analysis provides critical theoretical and practical insights into daily sales dynamics. The clear dominance of Month_Seq (*Bulan_ke*) with an importance score of 0.385 underscores that macro-level seasonal trends exert the strongest influence on unit demand in this retail context. This reflects broader consumer purchasing cycles and monthly promotional events, which cause substantial volume variations between months (e.g., from 391 units in February 2026 to 1,122 units in March 2026). Concurrently, *Lag_1* emerged as the second most influential predictor (0.294), confirming strong short-term auto-correlation in daily transaction records. This finding aligns with Della et al. [9], who identified previous-day sales as a key driver in short-term forecasting. While rolling window features (*Rolling_3* and *Rolling_7*) served to smooth out multi-day fluctuations, micro-temporal indicators such as *Is_Weekend* and *Day_of_Week* provided secondary refinements, capturing specific day-of-week purchasing behaviors.

From an operational standpoint, the proposed predictive model provides store management at Tulip Store MM Bekasi with a data-driven decision support tool:

- **Target Setting:** Replaces intuitive or static target allocation with realistic daily sales baselines based on recent sales momentum (*Lag_1*) and seasonal positioning (Month_Seq).
- **Inventory Allocation:** Enables store supervisors to forecast upcoming demand spikes, optimizing stock transfer from the central warehouse and mitigating both stockouts and holding costs for slow-moving articles.
- **Workforce Scheduling:** Assists store managers in aligning sales staff shift schedules with predicted high-volume days indicated by rolling sales trends and weekend indicators.

Despite these positive outcomes, several limitations must be acknowledged. First, the model was trained on a six-month dataset (October 2025 to March 2026), which may not fully capture long-term annual seasonality or major festive demand cycles (e.g., Eid al-Fitr or year-end sales). Second, exogenous variables—such as store-level discount percentages, mall foot traffic, external marketing campaigns, and competitor promotions—were not recorded in the POS dataset and thus excluded from feature engineering. Finally, the study focused exclusively on Random Forest Regression without cross-evaluating other advanced gradient boosting architectures like XGBoost, LightGBM, or CatBoost, which might yield further precision gains on time-series tabular data.

4 CONCLUSION

This study successfully developed a daily sales volume prediction model for women's undergarment products at Tulip Store MM Bekasi using Random Forest Regression. Based on six months of daily sales recapitulation data (October 2025 – March 2026), 11 engineered features spanning temporal, historical lag, and rolling statistics were utilized to predict daily unit sales (*Total_Qty*). Evaluation on 36 test data points demonstrated strong model accuracy, achieving a Mean Absolute Error (MAE) of 1.7649 and a Root Mean Square Error (RMSE) of 2.9740. Furthermore, 91.67% of predictions had errors within a tight margin of ± 5 units. Feature importance analysis identified *Bulan_ke* (0.385) and *Lag_1* (0.294) as the primary drivers of sales performance, confirming that macro-level monthly seasonality and previous-day sales momentum heavily dictate daily unit demand. These findings indicate that the model serves as an effective operational decision support tool for store management in setting data-backed daily targets, optimizing inventory replenishment, and managing shift schedules. Future research should expand the dataset timeframe to capture annual festive seasonality, incorporate external factors such as foot traffic and promotions, and compare Random Forest against other ensemble techniques like XGBoost or LightGBM.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Dr. Mukhlis, S.Kom., M.T., as the first supervisor, for providing guidance, direction, and valuable advice throughout the research process. The authors also thank Muhammad Yasir, S.Si., M.Kom., as the second supervisor, for his support and contributions in the writing and preparation of this manuscript. Additionally, the authors thank Tulip Store at Metropolitan Mall Bekasi and PT Star Asia Brothers for providing the sales data used in this study.

DECLARATIONS

Author Contributions:
Lusiana Situmorang: Conceptualization, Methodology, Software, Data curation, Writing–original draft.
Mukhlis: Supervision, Methodology, Writing–review & editing.

Funding:
The authors received no financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest:
The authors declare no conflict of interest regarding the publication of this paper.

Data Availability:
The data that support the findings of this study are available from the corresponding author upon reasonable request.

AI Use Statement:
During the preparation of this work, the authors used Claude AI in order to improve the language and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the final content of the published article.

Additional Information:
No additional information is available for this paper.

REFERENCES

[1] S. Anitha, "A Demand Forecasting Model Leveraging Machine Learning to Decode Customer Preferences for New Fashion Products," *Adv. Human-Computer Interact.*, vol. 2024, p. 8425058, 2024, doi: 10.1155/2024/8425058.

[2] M. Kunz, M. Koren, and S. Birr, “Deep Learning based Forecasting: a case study from the online fashion industry,” *arXiv preprint arXiv:2305.12345*, 2023.

[3] W. Januari, “Perancangan Sistem Informasi Penjualan Underwear Berbasis WEB Pada PT Star Asia Brothers Jakarta,” Skripsi, Universitas Nusa Mandiri, Jakarta, 2023.

[4] S. Falah, A. J. Tri, and N. Q. Nada, “Penerapan Algoritma Regresi Linier dan Random Forest Untuk Memprediksi Penjualan di Uchi Parfume,” *J. Teknol. Inf. dan Komput.*, vol. 7, no. 3, pp. 1317–1326, 2025.

[5] R. Hidayat et al., “Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket,” *J. Sains Komput. dan Inf.*, vol. 10, no. 1, pp. 101–109, 2025.

[6] M. S. Efendi, Sarwindo, and A. K. Zyen, “Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk,” *RESOLUSI: Rekayasa Teknol. Inf. dan Komput.*, vol. 5, no. 1, pp. 12–20, 2024, doi: 10.30865/resolusi.v5i1.2149.

[7] M. A. Kurniawan, G. Z. Syauqi, M. Safriyanti, F. U. Azmie, and A. Setiawan, “Prediksi Pendapatan Penjualan di Indomaret Menggunakan Algoritma Random Forest Regression,” *Pros. Sem. Nas. Komput. dan Teknol.*, pp. 93–99, 2025.

[8] R. Verdiyanto, D. Hartanti, and E. Purwanto, “Pengembangan Aplikasi Point of Sales untuk Prediksi Penjualan Harian Usaha Minuman Menggunakan Algoritma Random Forest Regression,” *J. Rekayasa Sist. Inf.*, vol. 8, no. 1, pp. 128–139, 2025.

[9] I. R. Della and D. Sunardi, “Analisis Kinerja Random Forest Dalam Memprediksi Penjualan Kantin Kasih Ibu Berdasarkan Data Historis,” *J. Inform. dan Teknol.*, vol. 8, pp. 371–379, 2026, doi: 10.51401/jit.v8i1.514.

[10] M. Sivakumar and S. Parthasarathy, “Trade-off between training and testing ratio in machine learning for medical image processing,” *PeerJ Comput. Sci.*, vol. 10, p. e2245, 2024, doi: 10.7717/peerj-cs.2245.

[11] S. Tripathi, D. Muhr, M. Brunner, H. Jodlbauer, M. Dehmer, and F. Emmert-Streib, “Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing,” *Front. Artif. Intell.*, vol. 4, p. 576892, 2021, doi: 10.3389/frai.2021.576892.

[12] J. N. Oruh and B. C. Iwuji, “A Hybrid Prediction Model for Classifying Student’s Academic Performance Using Voting Ensemble Method,” *Fudma J. Sci.*, vol. 10, no. 3, pp. 364–376, 2026, doi: 10.33003/fjs-2026-1003-4751.

[13] R. A. Casonatto, T. De Pádua Grillo Souza, and A. M. Mariano, “Quality and Risk Management in Data Mining: A CRISP-DM Perspective,” *Procedia Comput. Sci.*, vol. 242, pp. 161–168, 2024, doi: 10.1016/j.procs.2024.08.257.

[14] N. Lebkiri et al., “Using Machine Learning for Prediction Students Failure in Morocco: An Application of the CRISP-DM Methodology,” *Int. J. Educ. Inf. Technol.*, vol. 15, pp. 344–352, 2021, doi: 10.46300/9109.2021.15.36.

[15] M. Elkabalawy, A. Al-Sakkaf, E. M. Abdelkader, and G. Alfalah, “CRISP-DM-Based Data-Driven Approach for Building Energy Prediction Utilizing Indoor and Environmental Factors,” *Sustainability*, vol. 16, no. 17, p. 7249, 2024, doi: 10.3390/su16177249.

[16] M. Ricky and M. E. Al Rivan, “Implementasi Deep Convolutional Generative Adversarial Network untuk Pewarnaan Citra Grayscale,” *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 556–566, 2022, doi: 10.28932/jutisi.v8i3.5218.

[17] R. P. Nugroho, B. D. Setiawan, and M. T. Furqon, “Penerapan Metode Fuzzy Tsukamoto untuk Menentukan Harga Sewa Hotel (Studi Kasus: Gili Amor Boutique Resort, Dusun Gili Trawangan, Nusa Tenggara Barat),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2581–2588, 2019.

[18] Y. Pang et al., “Comparing multispectral and hyperspectral UAV data for detecting peatland vegetation patterns,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 132, p. 104043, 2024, doi: 10.1016/j.jag.2024.104043.

[19] M. Arabboev, S. Begmatov, M. Rikhsivoev, K. Nosirov, and S. Saydiakbarov, “A comprehensive review of image super-resolution metrics: classical and AI-based approaches,” *ACTA IMEKO*, vol. 13, no. 1, pp. 157–175, 2024, doi: 10.21014/actaimeko.v13i1.1679.

[20] K. Mahmoud et al., “Prediction of the effects of environmental factors towards COVID-19 outbreak using AI-based models,” *IAES Int. J. Artif. Intell.*, vol. 10, no. 1, pp. 35–42, 2021, doi: 10.11591/ijai.v10.i1.pp35-42.

[21] E. F. Rahayuningtyas, F. N. Rahayu, and Y. Azhar, “Prediksi Harga Rumah Menggunakan General Regression Neural Network,” *J. Inform.*, vol. 8, no. 1, pp. 59–66, 2021, doi: 10.31294/ji.v8i1.9036.

[22] K. Wang and X. Wang, “Logistics Transportation Vehicle Monitoring and Scheduling Based on the Internet of Things and Cloud Computing,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 8, pp. 50–61, 2024, doi: 10.14569/IJACSA.2024.0150806.