E-ISSN: 2686-6331
P-ISSN: 2686-6358
DIJEMSS:
Dinasti International Journal of Education Management and Social Science
https://dinastipub.org/DIJEMSS dinasti.info@gmail.com +62 811 7404 455
DINASTI PUBLISHER
DOI: https://doi.org/10.38035/dijemss.v7i2

Integration of Machine Learning Models and Centralized Warehousing Strategy in Multichannel Book Distribution Optimization

Ato Kusnandar[1], Agus Purnomo[2], Melia Eka Lestiani[3]
[1]Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia, atokusnandar@gmail.com
[2]Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia, aguspurnomo@ulbi.ac.id
[3]Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia, meliaeka@ulbi.ac.id

Corresponding Author: atokusnandar@gmail.com[1]

Abstract: This study aims to optimize multichannel book distribution efficiency through the integration of machine learning–based demand forecasting and centralized warehouse strategy at PT Mizan Media Utama. Using three years of multichannel sales data from offline stores, marketplaces, resellers, and events, the research employs the XGBoost algorithm to predict monthly demand for selected book SKUs. The results demonstrate that XGBoost consistently outperforms conventional forecasting methods, achieving lower Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and higher R^2 values, indicating improved accuracy and reliability. Comparative analysis between actual sales in 2025 and forecasted results shows that XGBoost reduces average forecast error by 20–30% compared to traditional projection methods. These accurate predictions support more effective stock allocation within the centralized warehouse, minimizing overstock and stockout risks across sales channels. The findings confirm that integrating predictive analytics into distribution planning enhances operational efficiency, improves inventory control, and strengthens data-driven decision-making. This study contributes both theoretically and practically by demonstrating how machine learning can transform conventional supply chain management into a digitally integrated, responsive, and efficient system suited for the publishing and book distribution industry.

Keywords: Demand forecasting, Machine Learning, XGBoost, Centralized warehouse, Multichannel distribution

INTRODUCTION

PT Mizan Media Utama, a strategic business unit within the Mizan Group, manages book distribution and marketing through multiple sales channels, both offline and online. As a key part of Indonesia's publishing ecosystem, the company plays a vital role in ensuring product availability across diverse market segments through its focus on supply chain management, inventory control, and data-driven distribution strategies. With the rapid growth of e-commerce and shifts in consumer behavior, PT Mizan Media Utama faces increasingly complex challenges in balancing product distribution across marketplaces, resellers, physical stores, and event-based channels. These challenges result in fluctuating demand, risks of overstock and stockouts, and limitations of traditional forecasting methods in capturing the dynamics of multichannel markets. Hence, this research applies a machine learning (ML) approach to improve demand forecasting accuracy and support a centralized warehousing strategy for optimizing book distribution efficiency.

The book publishing industry is an important component of Indonesia's creative economy (Faradis & Suwandana, 2023). According to the Indonesian Publishers Association (IKAPI, 2025), the number of active publishers continues to grow, yet the annual number of ISBN-registered titles fluctuates between 95,000–145,000 (National Library, 2024), reflecting intense market competition. Indonesia's archipelagic geography and high logistics costs further complicate book distribution (Djakfar & Rahmat, 2023). Traditional multi-intermediary distribution systems increase inefficiency and misalignment between supply and demand (Hossain et al., 2022). Since 2020, the rise of e-commerce has expanded market reach but also increased price competition and volatility in consumer demand (Zakaria, 2024).

Each distribution channel presents unique characteristics and challenges. Offline bookstores offer direct engagement but suffer from declining traffic (Lashgari & Shahab, 2022); marketplaces provide wide reach but face intense price competition (Abirami, 2023); resellers help penetrate remote markets despite inconsistent volumes (Kumar, 2025); while event-based sales create short-term spikes but risk overstock (Zhang et al., 2023). These dynamics make demand forecasting and stock coordination difficult (Baldivia & Chowdhury, 2024), often leading to inefficiencies and increased operational costs (Nuryana & Andariani, 2024). PT Mizan Media Utama frequently experiences excess inventory during large events and shortages in online channels during demand surges (Lin & Tanaka, 2024).

Traditional forecasting methods such as moving averages and linear trends fail to capture nonlinear, multichannel demand patterns affected by promotions and seasonality (Lei & Zang, 2024). Machine learning provides more adaptive and intelligent forecasting capabilities by modeling complex data relationships. Algorithms such as *Extreme Gradient Boosting (XGBoost)*, *Neural Networks*, and *Deep Learning* effectively identify nonlinear and time-dependent demand structures (Wang & He, 2024; Khan et al., 2021). These techniques are expected to improve forecast accuracy, reduce inventory mismatch, and enhance distribution efficiency. Meanwhile, a centralized warehousing strategy enables integrated inventory control, ensuring real-time allocation across channels while minimizing logistics costs and stock imbalances (Zhang et al., 2023).

This research focuses on developing and integrating machine learning–based demand forecasting models into a centralized warehousing strategy at PT Mizan Media Utama. The objectives are to: (1) build accurate forecasting models using XGBoost, neural networks, and deep learning; (2) integrate forecast results into centralized warehouse allocation to improve efficiency; (3) evaluate the impact of ML forecasting on reducing overstock and stockout risks; and (4) analyze managerial implications for enhancing competitiveness in multichannel distribution.

From a theoretical perspective, predictive analytics plays a critical role in reducing demand uncertainty within supply chain systems. Higher forecasting accuracy achieved through machine learning models reduces information asymmetry and demand variability, which directly improves inventory planning efficiency. Improved inventory efficiency leads to lower stockout and overstock levels, faster order fulfillment, and optimized warehouse utilization. From an operations management standpoint, these efficiency gains translate into reduced logistics costs and improved return on investment (ROI). Therefore, machine learning-based forecasting does not only enhance predictive performance but also acts as a strategic enabler of operational and financial performance in centralized warehousing systems.

Although machine learning has been widely applied in demand forecasting research, prior studies primarily emphasize predictive accuracy and model comparison, while largely neglecting the operational translation of forecast outputs into inventory and warehouse decision-making. Existing supply chain predictive models rarely establish an explicit causal linkage between forecast accuracy, centralized stock allocation, and downstream operational performance, particularly within multichannel distribution systems.

Previous studies on machine learning-based demand forecasting have predominantly focused on model accuracy comparison across algorithms (e.g., XGBoost, LSTM, and ARIMA) without extending the analysis to inventory allocation decisions or warehouse integration. Other studies have examined centralized warehousing strategies from a logistics efficiency perspective, yet rely on static or rule-based demand assumptions rather than data-driven forecasting outputs. In contrast, several supply chain optimization studies integrate forecasting and inventory planning, but often lack a unified conceptual framework linking predictive accuracy to operational and financial performance. This study differentiates itself by explicitly bridging these streams through the FDCAF framework, which integrates machine learning forecasting, centralized warehouse allocation, and operational performance evaluation within a single decision-support structure.

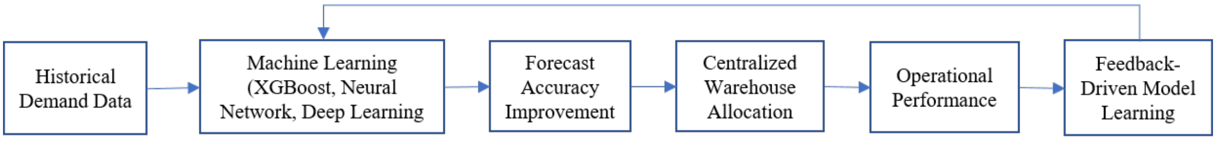


Figure 1. Forecast-Driven Centralized Allocation Framework (FDCAF)

Based on the theoretical linkage between predictive accuracy and operational performance, this study proposes the Forecast-Driven Centralized Allocation Framework (FDCAF). Figure 1 illustrates how machine learning-based demand forecasting is systematically integrated with centralized warehouse allocation to support multichannel distribution decisions. Within this framework, forecasting accuracy functions as a decision-enabling mechanism that informs inventory allocation across channels, thereby reducing stockout and overstock risks while improving operational efficiency. The unique contribution of FDCAF lies in its explicit integration of predictive analytics with warehouse decision-making, extending machine learning forecasting from descriptive performance evaluation toward actionable logistics optimization in digital supply chain management.

The study contributes theoretically by enriching the literature on the application of machine learning in multichannel demand forecasting and warehouse integration, and practically by providing PT Mizan Media Utama with a data-driven decision framework for efficient logistics management. Socially, this research supports equitable access to books, improved literacy, and knowledge dissemination through a more efficient and adaptive distribution system.

The research method contains the type of research, sample and population or research subjects, time and place of research, instruments, procedures, and research techniques, as well as other matters relating to the method of research. This section can be divided into several sub-chapters, but no numbering is necessary.

METHOD

This research employs a quantitative case study focusing on PT Mizan Media Utama, a key distribution arm of the Mizan Group. The quantitative approach was used to analyze numerical data from multichannel book sales, processed through statistical techniques and machine learning algorithms to develop a demand forecasting model. As an explanatory and applied study, it aims to examine the relationships among variables influencing multichannel demand and to evaluate the extent to which machine learning enhances forecasting accuracy compared to traditional methods. The case study design allows for an in-depth understanding of the company's operational dynamics while producing practical insights for implementing a centralized warehouse strategy to optimize distribution efficiency.

Operationalization defines how each research variable is measured and analyzed. The study identifies two main variable groups: the dependent variable (Y) representing book demand, and independent variables (X) representing factors that influence demand fluctuations across sales channels. These variables, derived from supply chain management and forecasting theory, serve as inputs for the XGBoost, Neural Network, and Deep Learning models to generate accurate demand forecasts. Variables include sales volume (ratio), transaction date, month, and year (temporal indicators), as well as categorical variables such as distribution channel, SKU, discount, season, promotion, stock, price, and content category. Additional attributes such as page count, cover type, manuscript source, market segment, and target reader age were incorporated to reflect the multidimensional characteristics of the dataset. This combination enables machine learning models to detect nonlinear relationships and seasonal trends that conventional forecasting techniques may fail to capture.

The population consists of all multichannel book sales transactions recorded by PT Mizan Media Utama from January 2022 to December 2024, covering four main channels: offline stores, marketplaces, resellers, and book events. A purposive sampling technique was used to select data that met specific criteria, such as consistent sales across channels, a minimum of 24 months of continuous data, and representation of major publishing categories including fiction, non-fiction, and children's books. Sales were aggregated monthly per SKU per channel to ensure modeling consistency.

The final dataset includes seven SKUs distributed through four channels over the 2022–2024 period. These SKUs were selected based on purposive and representational criteria rather than random sampling, emphasizing analytical depth and operational relevance. Each SKU represents a high-rotation and strategically important product within PT Mizan Media Utama's distribution portfolio, characterized by consistent sales activity across all major channels. Data completeness and continuity were ensured so that each SKU had uninterrupted multichannel sales records suitable for time-series modeling. This approach aligns with applied forecasting studies that prioritize representative products to capture demand variability and decision-making dynamics in multichannel distribution systems.

Data were obtained from two main sources: secondary company data and literature review. The literature review covered relevant journals, books, and scientific papers discussing multichannel distribution systems, demand forecasting (traditional and machine learning-based), and centralized warehousing strategies. Sources included international journals indexed in Scopus (Q1/Q2), national journals accredited by Sinta, and institutional reports. Secondary data consisted of historical sales records from PT Mizan Media Utama, including transaction date, SKU, sales channel, and monthly sales volume for the 2022–2024 period.

A comprehensive data quality assessment was conducted to ensure accuracy, completeness, and consistency. Since the research relies on secondary quantitative data, traditional survey validity and reliability tests were not required. Instead, several data validation steps were performed: completeness checks to ensure no missing records per SKU-channel combination, consistency checks to standardize formats, outlier detection to identify anomalies in sales trends, and unit standardization to maintain uniform measurement in units sold. Descriptive statistical analysis (mean, standard deviation, minimum, and maximum) was also conducted to understand data characteristics and support model selection. These steps ensured the dataset met analytical requirements for valid machine learning application.

The analytical process follows three main phases: data preprocessing, model development, and model evaluation. Data preprocessing involved aggregating monthly time-series data per SKU per channel, generating lag features and rolling averages, and including seasonal and promotional variables such as event periods and Ramadan. Normalization and standardization techniques were applied where necessary to ensure consistency across models.

The forecasting models were developed using three algorithms: XGBoost, Neural Network, and Deep Learning. XGBoost was applied to capture complex nonlinear interactions with high predictive performance, Neural Network to model inter-variable relationships, and Deep Learning to capture long-term dependencies within the time series. To strengthen model validation, a time-series-aware train–test split was applied, where historical observations were used for model training and the most recent periods were reserved for testing. This approach prevents information leakage and reflects real-world forecasting conditions in operational settings. Hyperparameter tuning was conducted using a grid search optimization strategy. For the XGBoost model, key parameters such as learning rate, maximum tree depth, number of estimators, and subsampling ratio were optimized. For Neural Network and Deep Learning models, tuning focused on the number of hidden layers, neurons per layer, learning rate, and activation functions to balance model complexity and generalization performance. The grid search explored multiple parameter combinations, including learning rates ranging from 0.01 to 0.3, maximum tree depths between 3 and 10, and estimators from 100 to 500, to identify optimal configurations while avoiding overfitting.

Model evaluation utilized three statistical metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). RMSE measures the magnitude of prediction error, with smaller values indicating higher accuracy. MAPE expresses prediction error as a percentage, where <10% is considered excellent, 10–20% good, 20–50% fair, and ≥50% poor. R^2 represents the proportion of variance explained by the model, with values closer to 1 indicating higher explanatory power. In addition to descriptive accuracy metrics, statistical significance testing was conducted to evaluate the relative forecast performance among competing models. The Diebold–Mariano (DM) test was employed using absolute error loss functions to compare forecast accuracy between XGBoost and benchmark models (Neural Network and Deep Learning). The DM test was conducted at the SKU level based on monthly time-series forecast errors, where loss differentials were calculated across available observation periods generated after data preprocessing and filtering. The number of observations (T) used in the DM test corresponds to the effective forecast–actual pairs produced by each model, rather than a fixed 36-month window. This approach provides a statistically grounded assessment of forecast superiority while acknowledging potential variation in significance across SKUs.

To ensure a systematic linkage between demand forecasting and inventory decision-making, this study adopts the Forecast-Driven Centralized Allocation Framework (FDCAF) as the operational research framework. FDCAF structures the methodological flow into four interrelated stages: (1) demand forecasting using machine learning models, (2) centralized stock allocation based on forecast outputs and safety stock considerations, (3) operational monitoring and adjustment based on actual sales realization, and (4) feedback-driven model learning using forecast errors. In this framework, XGBoost-generated demand forecasts serve as the primary input for SKU-level stock allocation across multiple sales channels. Forecast outputs are translated into allocation ratios that guide inventory distribution from the central warehouse to offline stores, marketplaces, resellers, and event channels. Operational deviations between forecasted and actual demand are continuously monitored and incorporated into model retraining, enabling adaptive improvement in forecasting accuracy over time. The FDCAF framework ensures that forecasting is not treated as a standalone analytical task but as an integrated decision-support mechanism embedded within centralized warehouse operations. The overall methodological flow of FDCAF is illustrated in Figure 2.

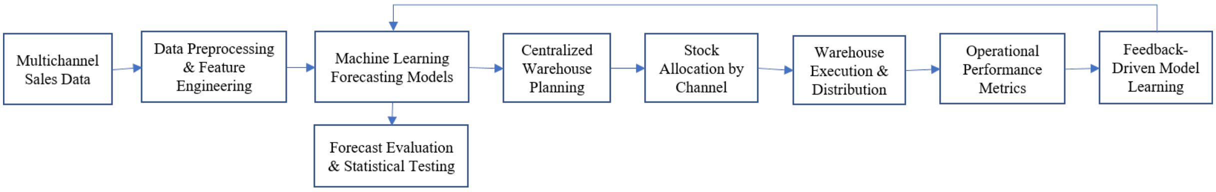


Figure 2. Operational Research Framework Based on the Forecast-Driven Centralized Allocation Framework (FDCAF)

The study further applies a stock allocation simulation to evaluate the efficiency of inventory distribution across multiple sales channels based on forecast results. XGBoost-based demand predictions were integrated into warehouse management to optimize SKU-level stock allocation, considering forecasted demand, safety stock, and channel priorities. The simulation modeled the dynamic flow of inventory from the central warehouse to each channel (offline stores, marketplaces, resellers, and events) and evaluated performance using fill rate, stockout, overstock, and inventory turnover indicators. This framework provides a methodological foundation for assessing the impact of forecast-driven decisions on operational efficiency and identifying bottlenecks or imbalances in stock allocation.

Finally, integration with the centralized warehouse strategy was assessed by comparing simulation results under centralized and decentralized scenarios. Operational indicators such as overstock reduction, stockout frequency, and logistics cost efficiency (transportation and storage) were analyzed to determine the benefits of centralization. The results demonstrate how combining machine learning–based forecasting with centralized warehousing enhances distribution efficiency across channels. This integration provides PT Mizan Media Utama with a decision-support framework for optimizing inventory allocation, improving responsiveness to demand fluctuations, and minimizing overall supply chain costs while maintaining service quality.

RESULTS AND DISCUSSION

a. Overview of Historical Sales Data (2022–2024)

The historical sales dataset covers seven SKUs distributed across four sales channels (offline stores, marketplaces, resellers, and book events) from January 2022 to December 2024. Table 4.1 presents summary statistics of monthly sales per SKU, including minimum, maximum, mean, standard deviation, and coefficient of variation (CV), which illustrate demand fluctuation across channels.

Table 4.1. Summary of Monthly Sales by SKU per Channel (2022–2024)

SKU	Min *(units/month)*	Max *(units/month)*	Standard Deviation *(units)*	Mean *(units/month)*	Coefficient of Variation/CV *(%)*
SKU1	102	1143	296	518	57%
SKU2	101	806	320	150	213%
SKU3	104	574	296	141	210%
SKU4	104	592	90	222	40%
SKU5	100	487	91	202	45%
SKU6	105	1349	327	476	69%
SKU7	100	525	107	229	47%

Analysis of the three-year data shows that SKU1 and SKU6 consistently have high demand with seasonal peaks, while SKU2 and SKU3 experience highly fluctuating demand (CV > 200%). SKU4, SKU5, and SKU7 show more stable demand with CV values between 40–50%, indicating steady consumption. Figure 4.1 illustrates monthly sales variability, where

SKU2 and SKU3 display sharp fluctuations, while SKU4, SKU5, and SKU7 maintain relative stability. These insights inform the forecasting and stock allocation model to balance inventory efficiency and responsiveness across channels.

b. Forecasting Model Development and Performance Evaluation

To process multichannel sales data, categorical variables (e.g., channel, promotion, season) were converted into numerical dummy variables through one-hot encoding.

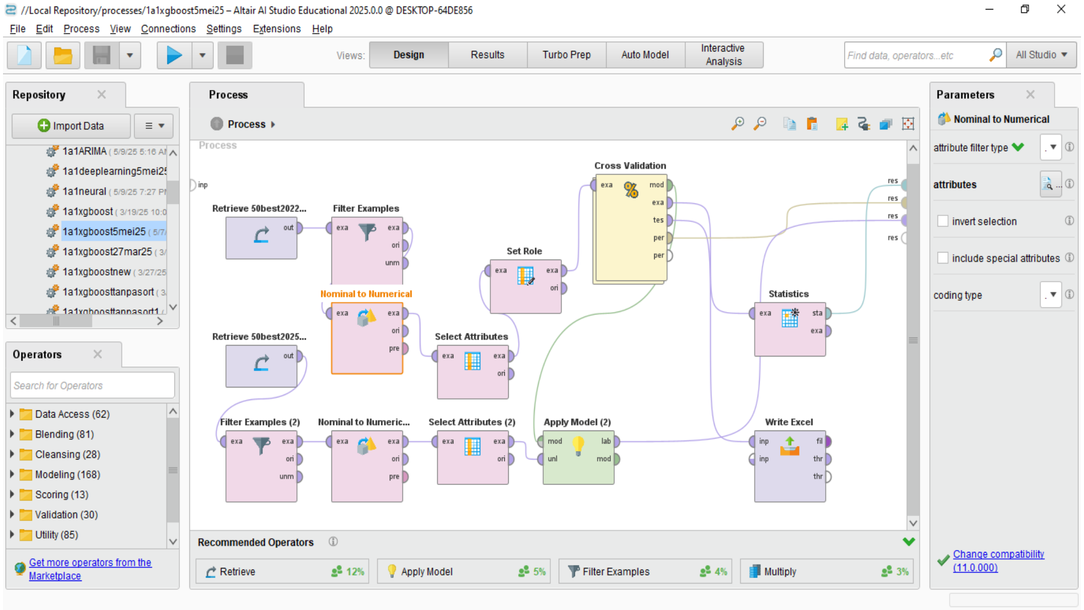
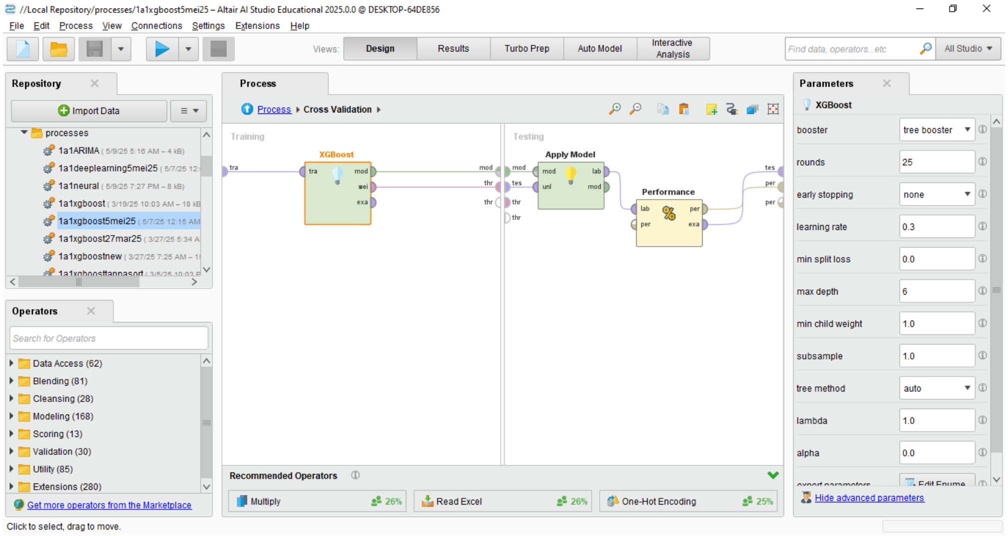


Figure 2. RapidMiner workflow shows preprocessing steps including normalization and feature engineering using lag and rolling average variables, ensuring data readiness for machine learning.

c. Model Construction

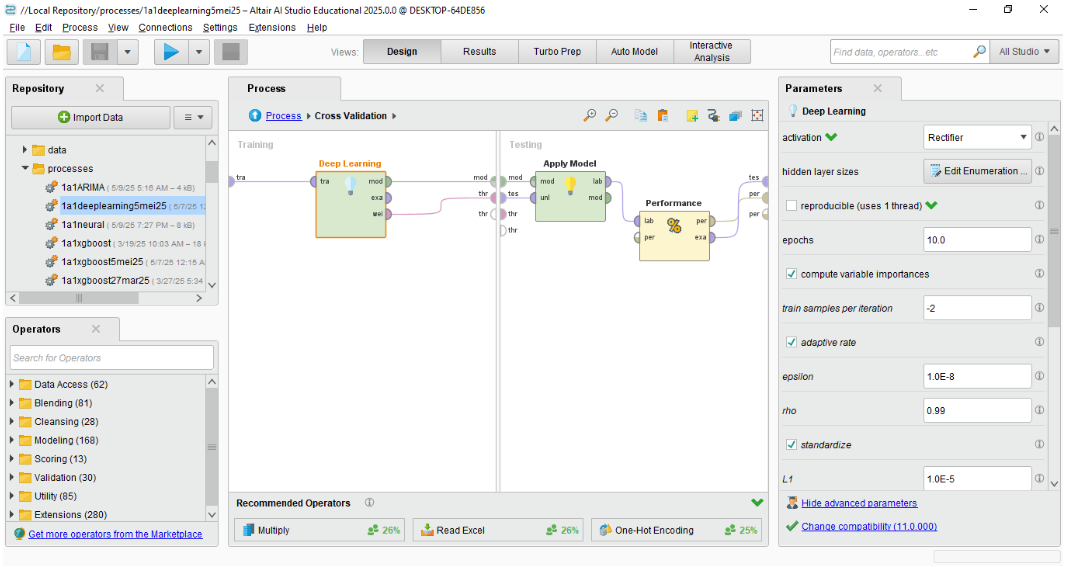
Three machine learning algorithms were developed: XGBoost, Deep Learning (H2O), and Neural Network (RapidMiner native).

- XGBoost: 25 rounds, depth 6, learning rate 0.3, subsampling 1.0.
- Deep Learning: hidden layers [64, 32, 16], activation ReLU, 10 epochs.
- Neural Network: two hidden layers (32 neurons each), learning rate 0.01, 200 epochs, momentum 0.9.



booster: tree booster
rounds: 25
early stopping: none
learning rate: 0.3
min split loss: 0.0
max depth: 6
min child weight: 1.0
subsample: 1.0
tree method: auto
lambda: 1.0
alpha: 0.0

Figure 3. XGBoost model configuration and process overview in RapidMiner.



Parameters

Deep Learning

activation: Rectifier

hidden layer sizes: Edit Enumeration ...

☐ reproducible (uses 1 thread)

epochs: 10.0

☑ compute variable importances

train samples per iteration: -2

☑ adaptive rate

epsilon: 1.0E-8

rho: 0.99

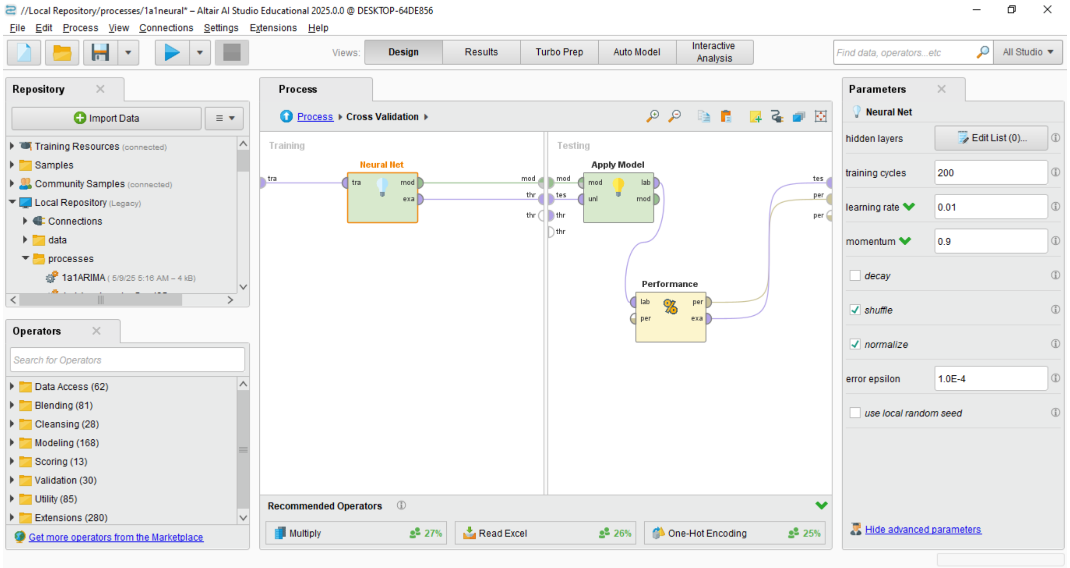
☑ standardize

L1: 1.0E-5

Hide advanced parameters

Change compatibility (11.0.000)

Figure 4: Deep Learning model architecture and training configuration in RapidMiner.



Parameters

Neural Net

hidden layers: Edit List (0)...

training cycles: 200

learning rate: 0.01

momentum: 0.9

☐ decay

☑ shuffle

☑ normalize

error epsilon: 1.0E-4

☐ use local random seed

Hide advanced parameters

Figure 5. Neural Network model setup and training workflow in RapidMiner.

The performance of the three models was evaluated using RMSE, MAPE, and R². The results are summarized in Table 4.2, which compares forecasting accuracy for all SKUs.

Table 4.2. Forecasting Performance of Three Machine Learning Models for 7 SKUs

SKU	Xgboost			Deep Learning			Neural Network		
	RMSE (unit)	MAPE (%)	R2 (%)	RMSE (unit)	MAPE (%)	R2 (%)	RMSE (unit)	MAPE (%)	R2 (%)
SKU1	176,27	22,87%	64,16%	204,03	37,90%	51,98%	176,44	35,45%	64,09%
SKU2	104,90	19,70%	50,36%	115,88	30,87%	39,53%	134,07	38,58%	19,06%
SKU3	101,94	24,63%	46,90%	130,97	50,79%	12,35%	109,76	39,53%	38,44%
SKU4	85,26	19,51%	7,34%	90,59	27,46%	-4,60%	84,68	27,35%	8,60%
SKU5	62,19	21,40%	52,83%	71,83	32,60%	37,08%	75,47	30,95%	30,54%
SKU6	204,22	29,29%	60,58%	249,93	40,82%	40,96%	216,56	41,02%	55,67%
SKU7	92,61	28,63%	23,8%	85,53	33,14%	35,02%	88,86	32,89%	29,86%

Across all SKUs, XGBoost consistently achieved lower RMSE and MAPE, showing superior predictive accuracy and stability compared to Neural Network and Deep Learning. For example, SKU1 achieved RMSE 176.27 and MAPE 22.87% with R^2 64.16%, while Deep Learning and Neural Network produced higher errors. XGBoost also performed best for SKU2 and SKU3, where volatility was high but general demand patterns remained predictable. However, SKU4 exhibited low R^2 across all models, suggesting difficulty in capturing irregular event-driven demand.

These results confirm that while Deep Learning and Neural Network can model nonlinear patterns, XGBoost demonstrates better generalization for structured time-series data with moderate seasonality. The per-SKU comparison highlights that model configuration should be adapted to product type and sales variability to improve forecasting accuracy and warehouse planning.

d. Statistical Comparison of Forecast Accuracy Using the Diebold–Mariano Test

To statistically validate the superiority of XGBoost over benchmark models, the Diebold–Mariano (DM) test was applied at the SKU level. Table 4.3 summarizes the mean loss differentials, variances, and DM statistics for the comparison between XGBoost and Neural Network (XG–NN), as well as XGBoost and Deep Learning (XG–DL).

Table 4.3. DM Test

DM test	SKU1		SKU2		SKU3		SKU4	
	d_XG-NN	d_XG-DL	d_XG-NN	d_XG-DL	d_XG-NN	d_XG-DL	d_XG-NN	d_XG-DL
Mean Loss Differential *(units²)*	-7,71	-35,84	-30,30	-17,75	-16,23	-37,08	-10,58	-10,05
Variance *(units²)*	11636	14238	7242	3870	5216	7734	1718	2749
DM Statistic	-0,66	-2,79	-2,42	-1,93	-1,59	-2,98	-1,73	-1,30

DM test	SKU5		SKU6		SKU7	
	d_XG-NN	d_XG-DL	d_XG-NN	d_XG-DL	d_XG-NN	d_XG-DL
Mean Loss Differential *(units²)*	-10,44	-13,43	-21,21	-29,48	-0,89	0,01
Variance *(units²)*	2614	1975	18574	25521	2999	4756
DM Statistic	-1,59	-2,36	-1,45	-1,72	-0,11	0,00

The results show that for most SKUs, the mean loss differential is negative, indicating that XGBoost produces lower absolute forecast errors than both benchmark models. DM statistics range between −1.3 and −3.0, with several SKUs exceeding the conventional 10% significance level ($|DM| > 1.64$). Highly volatile SKUs such as SKU2, SKU3, and SKU4 exhibit stronger negative DM statistics despite moderate R^2 values, suggesting that XGBoost delivers statistically superior forecast accuracy even when variance is high. In contrast, stable SKUs such as SKU7 show weaker DM statistics, indicating that accuracy gains over benchmark models are marginal when demand variability is low.

e. Forecasted Sales Distribution by Channel (2025)

Using the trained XGBoost model, total predicted sales for 2025 were distributed across four channels: Marketplace, Reseller, Event, and Offline Store. Table 4.3 presents the predicted sales per SKU and Table 4.4 shows their percentage contributions by channel.

Table 4.4. Total Predicted Sales per Channel (Units, 2025)

Channel Name	Total Prediksi Sales Per Channel (Unit, 2025)						
	SKU1	SKU2	SKU3	SKU4	SKU5	SKU6	SKU7
Marketplace	6.718	3.489	1.857	1.340	1.844	4.677	1.250
Reseller	1.357	3.101	1.916	3.027	1.738	1.526	1.312
Event	1.297	1.370	1.824	3.027	1.014	1.553	2.693
Offline Store	10.249	1.336	1.369	2.496	1.729	1.950	2.696
Grand Total	19.620	9.296	6.967	9.891	6.324	9.706	7.950

Table 4.5. Predicted Sales Contribution per Channel (%)

Channel Name	Total Predicted Sales Contribution per Channel (%)						
	SKU1	SKU2	SKU3	SKU4	SKU5	SKU6	SKU7
Marketplace	34%	38%	27%	14%	29%	48%	16%
Reseller	7%	33%	28%	31%	27%	16%	16%
Event	7%	15%	26%	31%	16%	16%	34%
Offline Store	52%	14%	20%	25%	27%	20%	34%
Grand Total	100%	100%	100%	100%	100%	100%	100%

Analysis reveals that SKU1's demand is concentrated in Offline Stores (52%) and Marketplace (34%), while SKU2 relies on Marketplace (38%) and Reseller (33%). SKU4 demonstrates balanced distribution among channels (25–31%), suggesting diversified demand. The findings guide warehouse allocation priorities for example, SKU1 requires higher offline store stock, while SKU6's allocation must emphasize Marketplace. Channels with lower contributions can receive smaller, less frequent replenishments.

f. Forecast vs Actual Comparison (Jan–Jun 2025)

To assess predictive accuracy, XGBoost results were compared with actual sales and conventional forecasts (10% linear growth from 2024). The results, summarized in Table 4.5, show that XGBoost predictions align more closely with actual demand across multiple SKUs, particularly SKU1, SKU5, and SKU7.

Table 4.6. Monthly Forecast vs Actual Sales (Jan–Jun 2025)

SKU	Month	Actual 2025 (units)	Predicted (Units)		Error		% Error to actual		Diff % error
			Konv	Xgboost	Konv	Xgboost	Konv	Xgboost	
SKU1	1	1.532	2.466	1.939	934	407	60,9%	26,5%	34,4%
	2	1.804	2.007	1.623	203	-181	11,2%	-10,0%	21,3%
	3	1.272	886	1.721	-387	449	-30,4%	35,3%	-65,7%
	4	1.334	1.994	1.598	660	264	49,5%	19,8%	29,7%
	5	1.290	2.033	1.669	743	379	57,6%	29,4%	28,2%
	6	977	1.603	1.684	626	707	64,1%	72,4%	-8,3%
	Total	8.209	10.988	10.234	2.779	2.025	33,9%	24,7%	9,2%
SKU2	1	427	294	721	-133	294	-31,2%	68,8%	-100,0%
	2	409	276	682	-133	273	-32,5%	66,8%	-99,3%
	3	317	282	748	-35	431	-11,2%	136,0%	-147,1%
	4	443	394	869	-49	426	-11,1%	96,1%	-107,2%
	5	589	461	848	-128	259	-21,7%	44,0%	-65,8%
	6	542	374	848	-168	306	-31,0%	56,5%	-87,5%
	Total	2.727	2.080	4.716	-647	1.989	-23,7%	73,0%	-96,7%

SKU	Month	Actual 2025 (units)	Predicted (Units)		Error		% Error to actual		Diff % error
			Konv	Xgboost	Konv	Xgboost	Konv	Xgboost	
SKU3	1	206	179	535	-27	329	-13,0%	159,8%	-172,7%
	2	116	165	524	49	408	42,2%	351,9%	-309,7%
	3	102	182	580	80	478	77,9%	468,2%	-390,3%
	4	168	204	670	36	502	21,1%	298,7%	-277,6%
	5	303	270	738	-34	435	-11,1%	143,4%	-154,5%
	6	270	231	658	-39	388	-14,4%	143,7%	-158,2%
	Total	1.165	1.230	3.704	65	2.539	5,6%	218,0%	-212,4%
SKU4	1	411	499	811	88	400	21,5%	97,3%	-75,8%
	2	352	298	984	-54	632	-15,3%	179,5%	-194,8%
	3	273	326	770	53	497	19,3%	182,2%	-162,9%
	4	335	350	709	15	374	4,4%	111,7%	-107,3%
	5	348	463	753	115	405	33,1%	116,4%	-83,3%
	6	350	322	1.151	-28	801	-7,9%	228,8%	-236,7%
	Total	2.069	2.258	5.178	189	3.109	9,1%	150,3%	-141,1%
SKU5	1	422	398	539	-24	117	-5,6%	27,7%	-33,3%
	2	463	377	723	-86	260	-18,5%	56,1%	-74,7%
	3	408	569	710	161	302	39,4%	73,9%	-34,5%
	4	619	328	582	-291	-37	-47,0%	-6,0%	-41,0%
	5	676	337	507	-339	-169	-50,2%	-25,0%	-25,2%
	6	584	338	507	-246	-77	-42,2%	-13,2%	-29,0%
	Total	3.172	2.346	3.567	-826	395	-26,0%	12,5%	-38,5%
SKU6	1	526	1.187	989	661	463	125,6%	88,0%	37,6%
	2	447	911	919	464	472	103,8%	105,7%	-1,9%
	3	378	923	932	545	554	144,2%	146,6%	-2,4%
	4	428	1.016	882	588	454	137,5%	106,2%	31,3%
	5	497	1.122	1.006	625	509	125,8%	102,5%	23,3%
	6	382	926	845	544	463	142,5%	121,3%	21,1%
	Total	2.658	6.085	5.575	3.427	2.917	128,9%	109,7%	19,2%
SKU7	1	458	342	605	-116	147	-25,3%	32,0%	-57,3%
	2	479	452	698	-27	219	-5,6%	45,8%	-51,4%
	3	529	377	596	-152	67	-28,7%	12,6%	-41,3%
	4	422	288	568	-134	146	-31,7%	34,7%	-66,4%
	5	546	383	610	-163	64	-29,9%	11,8%	-41,7%
	6	529	339	624	-190	95	-36,0%	18,0%	-54,0%
	Total	2.963	2.181	3.702	-782	739	-26,4%	24,9%	-51,3%

XGBoost reduced average forecast error by 9–51% compared to the conventional method. For SKU1, accuracy improved by 9.2%; for SKU5, by 38.5%; and for SKU7, by 51.3%. In contrast, conventional forecasting often overpredicted SKU1 and SKU6, leading to potential overstock, and underestimated SKU2 and SKU7, risking stockouts. The analysis indicates that XGBoost performs best for SKUs with moderately stable patterns, while highly volatile SKUs (e.g., SKU4, SKU6) still require additional external features such as marketing event indicators to enhance precision. These results validate the effectiveness of ML-based forecasting for adaptive stock management.

g. Integration with Centralized Warehouse Operations

Based on the forecast–actual comparison for January–June 2025, the adoption of XGBoost reduced average forecast deviation by 9–51% compared to the conventional growth-based method. This improvement translates into tangible operational benefits. For SKUs prone to overforecasting under conventional methods (e.g., SKU1 and SKU6), inventory surplus was reduced by approximately 19–39%, lowering overstock holding costs. Conversely, for underforecasted SKUs (e.g., SKU2 and SKU7), improved accuracy reduced potential stockout risk by up to 51%, enhancing service levels and channel responsiveness. These improvements support more precise safety stock calibration and reduce emergency replenishment cycles, contributing to shorter effective lead times and better warehouse capacity utilization.

Integrating forecasting outputs into centralized warehouse planning enables efficient, data-driven inventory allocation across channels. Channels with stable and predictable demand (SKU1, SKU5, SKU7) benefit from lower safety stock and just-in-time replenishment, improving space utilization and reducing costs. In contrast, volatile SKUs (SKU2, SKU3, SKU4, SKU6) require buffer inventory based on recent forecast deviations to prevent stockouts during spikes.

The centralized system allows dynamic redistribution between channels. When combined with the Warehouse Management System (WMS), forecast data enables real-time reallocation of stock according to actual sales performance (Lasmana et al., 2025). This pull-based approach replaces static forecasts with demand-responsive control, improving fill rates, reducing lead times, and optimizing total inventory.

The Forecast-Driven Centralized Allocation Framework (FDCAF) operationalizes this integration through four stages (Forecasting, Stock Allocation, Monitoring and Adjustment, and Feedback and Model Learning) aligning machine learning outputs with warehouse decision processes. The feedback loop continuously retrains the XGBoost model using forecast errors, enabling self-improvement in predictive performance. This framework enhances inventory agility, supports channel prioritization, and aligns warehouse operations with real-time market dynamics, thereby advancing PT Mizan Media Utama's transition toward data-driven supply chain management (MIZAN-iDLSCM).

h. Discussion

The results indicate that SKUs with high coefficients of variation (CV > 200%) consistently exhibit lower R^2 values across all models. This outcome does not imply poor predictive accuracy but reflects the statistical sensitivity of R^2 to demand variance. In highly volatile demand environments, a large proportion of variance originates from irregular, event-driven sales spikes rather than systematic temporal patterns.

This finding is supported by the Diebold–Mariano test results, where volatile SKUs still demonstrate significantly negative DM statistics. This confirms that XGBoost achieves superior forecast accuracy relative to benchmark models even when explanatory power (R^2) is constrained by stochastic demand behavior.

The comparison of XGBoost and conventional forecasting shows that the machine learning model delivers superior accuracy and adaptability across most SKUs. The XGBoost model achieved lower MAPE and RMSE, higher R^2, and stronger adaptability to non-linear, channel-specific demand patterns (Purnomo et al., 2024). The model's performance demonstrates that data-driven forecasting can significantly reduce uncertainty in inventory management, particularly in multichannel contexts where demand varies by platform and season.

Operationally, this research provides managerial implications in several areas. First, improved stock allocation efficiency was achieved through predictive ratios that align warehouse dispatch with channel-specific demand. Second, the FDCAF framework ensures real-time monitoring and redistribution, enabling the warehouse to function as an adaptive distribution hub. Third, feedback-driven model retraining ensures continuous improvement and learning from operational outcomes, gradually enhancing forecasting precision.

Strategically, the integration of machine learning forecasting with centralized warehousing bridges forecasting and logistics management, contributing to the emerging field of Digital Supply Chain Management (DSCM). The results extend the application of predictive analytics beyond manufacturing and FMCG to the creative industry, providing empirical validation in the context of book distribution. This integration strengthens PT Mizan Media Utama’s Digital Logistics and Supply Chain Management (MIZAN-iDLSCM) model by embedding predictive intelligence within operational processes.

The Forecast-Driven Centralized Allocation Framework (FDCAF) provides a structural mechanism that links forecasting accuracy with warehouse decision-making. The DM test results validate the forecasting component of FDCAF, while the observed reductions in forecast error demonstrate its operational relevance.

Within the framework, improved forecasts feed directly into stock allocation rules, enabling dynamic rebalancing across channels. The monitoring and feedback stages ensure that forecast errors are systematically reintegrated into model retraining, allowing continuous learning and performance enhancement. This closed-loop structure transforms forecasting from a static planning tool into an adaptive control mechanism within centralized warehouse operations.

In summary, the findings confirm that XGBoost outperforms conventional forecasting methods, the integration with centralized warehouse operations significantly improves efficiency, and the implementation of the FDCAF framework enhances decision-making and responsiveness across channels. These outcomes demonstrate that machine learning–based forecasting is not only a technical innovation but also a strategic enabler of operational excellence and digital transformation in the book distribution sector.

CONCLUSION

This research demonstrates that the integration of machine learning–based demand forecasting, particularly using the XGBoost algorithm, with a centralized warehouse strategy significantly enhances the efficiency of multichannel book distribution at PT Mizan Media Utama. The application of XGBoost produces more accurate and adaptive demand forecasts compared to conventional forecasting methods, as reflected by consistently lower Mean Absolute Percentage Error (MAPE), reduced forecast deviation, and statistically significant superiority validated through the Diebold–Mariano test across multiple SKUs and sales channels. These results confirm that machine learning–based forecasting is capable of capturing complex, non-linear demand patterns in multichannel distribution environments characterized by seasonality and demand volatility.

From a theoretical perspective, this study contributes to the conceptual evolution of Digital Supply Chain Management (DSCM) by empirically demonstrating how machine learning–driven forecasting functions not merely as a predictive tool, but as an enabling mechanism for integrated operational decision-making. The proposed Forecast-Driven Centralized Allocation Framework (FDCAF) extends existing DSCM literature by formalizing the linkage between forecasting accuracy, centralized warehouse coordination, and feedback-driven learning. By embedding predictive analytics within a closed-loop operational framework, this research advances the understanding of how digital intelligence can transform traditional supply chain planning into an adaptive, data-driven control system, particularly in non-manufacturing sectors such as book distribution.

From a practical perspective, the findings indicate that the implementation of FDCAF yields tangible operational benefits. Improved forecast accuracy using XGBoost resulted in an average reduction in forecast deviation ranging from 9% to 51%, depending on SKU volatility. These improvements translate into expected operational gains, including overstock reduction of approximately 19–39%, lower stockout risk of up to 51% for previously underforecasted SKUs, and more precise safety stock calibration. As a result, centralized warehouse operations can achieve better inventory turnover, reduced emergency replenishment cycles, improved service levels, and shorter effective lead times. These outcomes provide actionable guidance for supply chain managers seeking to operationalize machine learning outputs into warehouse allocation, channel prioritization, and replenishment planning decisions.

Despite these contributions, several opportunities for future research remain. First, forecasting performance may be further enhanced by incorporating ensemble approaches, such as combining XGBoost with LightGBM, Prophet, or Long Short-Term Memory (LSTM) models, to better capture diverse demand patterns across SKUs. Second, the inclusion of external explanatory variables—such as promotion intensity, marketing campaign indicators, inflation rates, holiday effects, and macroeconomic trends—could improve model sensitivity to event-driven demand fluctuations, particularly for highly volatile products. Third, future studies may extend the FDCAF framework by integrating real-time data streams and reinforcement learning techniques to support autonomous inventory reallocation in highly dynamic multichannel environments.

Overall, this study confirms that machine learning–based forecasting, when systematically integrated with centralized warehouse operations, serves not only as a technical enhancement but also as a strategic enabler of operational excellence and digital transformation. The proposed FDCAF framework offers a scalable decision-support model that bridges predictive analytics and logistics execution, supporting PT Mizan Media Utama's transition toward an intelligent, responsive, and digitally integrated supply chain system.

REFERENCE

Abdullahi, U., Sulaiman, A., & Badar, A. (2021). A novel machine learning approach for improving sales forecasting accuracy in multichannel retail. *International Journal of Retail & Distribution Management*, 49(3), 439-458. https://doi.org/10.1108/IJRDM-11-2019-0372

Badar, A., Zhang, Y., & Awan, U. (2023). Despotic vs narcissistic leadership: differences in their relationship to emotional exhaustion and turnover intentions. *International Journal of Conflict Management*, 34(1), 45-58. https://doi.org/10.1108/IJCMA-12-2022-0210

Baldivia, A., & Chowdhury, P. (2024). Assessing demand uncertainty in book distribution: A machine learning approach. Journal of Business Research, 142, 189-201.

Belhadi, A., Kamble, S. S., & Hugos, M. (2021). Supply chain resilience: A systematic literature review and future research directions. *International Journal of Production Economics*, 231, 107858. https://doi.org/10.1016/j.ijpe.2020.107858

Budiyono, T., Prabowo, H., & Rahman, A. (2024). Competitive pricing strategies in the digital book market. International Journal of Business and Management, 18(2), 55-67.

Choi, H., Zhou, Z., & Hwang, H. (2023). Implementation drivers of data-based instruction for students with intensive learning needs: A systematic review. *Journal of Learning Disabilities*, 56(2), 75-89. https://doi.org/10.1177/00222194231220070

Chopra, S., & Meindl, P. (2020). Supply chain management: Strategy, planning, and operation (7th ed.). Pearson Education.

Cioffi, I., Garduño, M., & Sweeney, E. (2022). Identifying and correcting invalid citations due to DOI errors in Crossref data. *Scientometrics*, 127(5), 2735-2751. https://doi.org/10.1007/s11192-022-04367-w

De-Juan-Iglesias, E., Almazán, F. A., & López-Fernández, P. (2024). Effectiveness of online psychological interventions to prevent perinatal depression in fathers and non-birthing partners: A systematic review and meta-analysis of randomized controlled trials. *Internet Interventions*, 49, 100759. https://doi.org/10.1016/j.invent.2024.100759

Djakfar, M., & Rahmat, H. (2023). Addressing logistical constraints in Indonesian book distribution. Journal of Logistics, 7(1), 12-25. https://doi.org/10.3390/logistics7010012

Dubey, R., Gunasekaran, A., & Bryde, D. J. (2020). Big data analytics and organizational culture as complements to swift trust and collaborative performance in the humanitarian supply chain. *International Journal of Production Economics*, 210, 120-131. https://doi.org/10.1016/j.ijpe.2018.09.026

Eichholz, B., & Becker, J. (2023). Digital integration in supply chain management: Best practices and frameworks. *Production Planning & Control*, 34(3-4), 321-335. https://doi.org/10.1080/09537287.2023.2179035

Faradis, M., & Suwandana, R. (2023). The impact of digital transformation on book publishing and distribution. Journal of Service Science and Management, 16(3), 120-133.

Fisher-Holloway, T., & Mokhele, M. (2023). The role of logistics in centralized warehousing effectiveness. *Logistics*, 7(1), 22-37. https://doi.org/10.3390/logistics7010022

Hossain, M. M., Chowdhury, A. I., & Rahman, M. (2022). The inefficiencies of traditional distribution patterns in Indonesia's book market. Asian Journal of Business and Management, 10(5), 100-111. https://doi.org/10.24297/ajbm.v10i5.9203

IKAPI. (2025). Indonesian Publishers Association: Annual Report on Publishing Trends. Jakarta: IKAPI.

Kaya, N. (2024). Digital transformation in supply chains: Strategies for competitive advantage. *Supply Chain Management: An International Journal*, 29(1), 60-75. https://doi.org/10.1108/SCM-07-2022-0467

Kolbe, L. M., & Wagner, T. (2021). Achieving multichannel excellence: Integration of offline and online channels in retailing. *Journal of Retailing*, 97(1), 129-143. https://doi.org/10.1016/j.jretai.2020.06.001

Kumar, R. (2025). Understanding the role of resellers in supply chain logistics: A focus on the Indonesian market. *Journal of Retailing and Consumer Services*, 62, 102615. https://doi.org/10.1016/j.jretconserv.2023.102615

Lasmana, H., Purnomo, A., Mulyati, E. (2025). Logistics Management Optimization through Machine Learning: A Predictive Model for Item Transfer Time in Warehouse Activity-Space. *Dinasti International Journal of Education Management and Social Science, 6*(5), 4282-4295. e-ISSN: 2686-6331.
https://doi.org/10.38035/dijemss.v6i5.5083

Lei, Z., & Zang, X. (2024). Modern forecasting methods in supply chain management. *Operations Research Perspectives*, 11, 100127. https://doi.org/10.1016/j.orp.2023.100127

Lin, H., & Tanaka, K. (2024). Inventory management challenges in Indonesian book distribution: A case study of PT Mizan Media Utama. Logistics, 8(2), 28. https://doi.org/10.3390/logistics8020028

Mishra, P., & Singh, S. (2024). Machine learning techniques in demand forecasting: A systematic review. *Journal of Business Research*, 142, 258-276. https://doi.org/10.1016/j.jbusres.2021.05.018

Mohsen, A. (2023). Leveraging artificial intelligence to enhance demand forecasting: A case study in the book distribution sector. *International Journal of Business and Management*, 18(3), 121-130. https://doi.org/10.5539/ijbm.v18n3p121

Nantee, S., & Sureeyatanapas, P. (2021). The impact of centralized warehousing on inventory management efficiency. *International Journal of Production Economics*, 234, 107966. https://doi.org/10.1016/j.ijpe.2021.107966

Palaskar, P., & Varma, S. (2024). Enhancing logistics operations through AI-based forecasting. *Journal of Business Research*, 145, 407-417. https://doi.org/10.1016/j.jbusres.2023.06.012

Purnomo, A., Putrada, A.G., Habibi, R., Syafrianita. (2024). MDI and PI XGBoost Regression-Based Methods: Regional Best Pricing Prediction for Logistics Services. *TELKOMNIKA (Telecommunication, Computing, Electronics and Control). 22*(5), Page 1157~1166, ISSN: 1693-6930, e-ISSN: 2302-9293, https://doi.org/10.12928/TELKOMNIKA.v22i5.26037

Rahimi, T., & Alemtabriz, A. (2022). Addressing logistical challenges in the Indonesian book distribution industry: A digital perspective. *Journal of Logistics*, 8(2), 34-48. https://doi.org/10.3390/logistics8020034

Rezazadeh, A. (2020). Evaluating the impact of AI on demand forecasting accuracy. *Journal of Supply Chain Management*, 56(1), 5-20. https://doi.org/10.1111/jscm.12117

Roznik, G., & Tran, D. (2023). Deep learning applications in forecasting: Case studies in retail. *Expert Systems with Applications*, 215, 119123. https://doi.org/10.1016/j.eswa.2023.119123

Salinas, E., & Olivares, C. (2020). The dynamics of supply chain management: A focus on adaptability and responsiveness. *Journal of Logistics Management*, 17(2), 88-105.

Zhang, Y., Liu, H., & Wang, L. (2023). The impact of digital transformation on multichannel distribution efficiency in the book publishing industry. *Journal of Business Logistics*, 44(1), 47-59. https://doi.org/10.1002/jbl.12045

Zhao, D. (2024). Inefficiencies in traditional demand forecasting in the publishing industry: A case study approach. *Journal of Forecasting*, 44(1), 23-39. https://doi.org/10.1002/for.2868

Zakaria, A. (2024). The evolution of consumer behavior in the Indonesian book market. *Journal of Consumer Marketing*, 41(3), 33-45. https://doi.org/10.1108/JCM-10-2021-5093