EVALUATING THE IMPACT OF HYPERPARAMETER TUNING ON THE PERFORMANCE OF AN ESEMBLE RF-SVM MODEL FOR SME SALES PREDICTION

Eko Purwanto[1*], Bangun Prajadi Cipto Utomo[2], Hanifah Permatasari[3], Farahwahida Mohd[4]
Universitas Duta Bangsa[1,2,3], Universiti Kuala Lumpur[4]
*Correspondence Email : eko_purwanto@udb.ac.id

ABSTRACT

Accurate demand forecasting is a critical enabler of digital transformation for fashion-retail Small and Medium Enterprises (SMEs), whose sales patterns are strongly shaped by seasonality, promotions, and rapidly shifting trends. Prior hybrid Random Forest–Support Vector Machine (RF-SVM) models have reported near-perfect accuracy in predicting retail transaction value, but were trained with default hyperparameters and were not evaluated on demand-volume forecasting. This study evaluates the extent to which systematic hyperparameter tuning improves the performance of an ensemble RF-SVM model when applied to weekly demand-volume prediction for fashion-retail SMEs. Weekly sales records from retail SMEs in the Surakarta region (Indonesia) were collected, pre-processed, and split into training and test sets (80:20). Random Forest and Support Vector Regression base learners were each optimised through Grid-Search cross-validation, and their tuned predictions were combined using a validation-optimised weighted hybridisation scheme. The tuned ensemble RF-SVM model achieved a Mean Squared Error (MSE) of 0.001109, a Root Mean Squared Error (RMSE) of 0.033308, and a coefficient of determination (R^2) of 0.981786 on normalised test data. Predicted-versus-actual and residual plots indicate close tracking of actual demand with errors randomly distributed around zero, showing no systematic bias. Explicit hyperparameter tuning meaningfully strengthens the robustness of ensemble RF-SVM models against SME demand fluctuations, offering a practical route to more reliable, data-driven inventory and promotion planning for retail SMEs undergoing digital transformation.

KEYWORDS Hyperparameter Tuning, Ensemble Learning, Random Forest, Support Vector Machine, Demand Forecasting, Small and Medium Enterprises

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International

INTRODUCTION

Digital transformation has pushed Small and Medium Enterprises (SMEs) across sectors to adopt data-driven technologies to strengthen competitiveness and operational efficiency. Among fashion-retail SMEs in particular, demand uncertainty—driven by seasonality, changing fashion trends, promotional activity, and volatile consumer behaviour—remains a central operational challenge (Bhavsar et al., 2025; Chabalala et al., 2024). Inaccurate demand forecasts translate directly into excess inventory, stockouts, and inflated operating costs.

Conventional forecasting techniques, such as linear regression and classical time-series analysis, struggle to capture the non-linear, multi-dimensional relationships that characterise modern SME sales data. Consequently, machine learning approaches—particularly hybrid models that combine complementary algorithms—have gained increasing attention for improving forecast accuracy (Chen et al., 2025; Warburton, 2025).

Prior work has shown that combining Random Forest (RF) and Support Vector Machine (SVM) can outperform single-algorithm approaches, since RF handles high-dimensional data and mitigates overfitting while SVM captures non-linear decision boundaries (Ganguly & Mukherjee, 2024). However, such hybrid RF-SVM approaches have predominantly been applied to predicting retail transaction value using default hyperparameters, leaving two important gaps: (1) the extent to which systematic hyperparameter tuning—rather than default configurations—affects hybrid model performance remains largely unquantified, and (2) demand-volume forecasting at the weekly level for fashion-retail SMEs, which exhibits more pronounced seasonal and trend-driven fluctuation than general retail transaction data, has received comparatively little attention (M.V.N, 2025).

Hyperparameter tuning—commonly performed via Grid Search or Bayesian optimisation—has been widely shown to improve machine learning model accuracy by systematically identifying parameter configurations best suited to the data at hand (Lai, 2024; Yuvaraj, 2025). Yet its incremental contribution when layered onto an already-hybridised RF-SVM model has not been directly evaluated in the SME forecasting literature.

This study addresses that gap by evaluating the impact of explicit, systematic hyperparameter tuning on the performance of an ensemble RF-SVM model applied to weekly demand-volume prediction for fashion-retail SMEs. The specific objectives are to: (1) develop a tuned ensemble RF-SVM model capable of capturing non-linear and seasonal patterns in SME demand data; (2) empirically assess the contribution of hyperparameter tuning relative to default-parameter baselines reported in the literature; and (3) derive practical implications for SMEs seeking to strengthen data-driven decision-making as part of their digital transformation.

RESEARCH METHOD

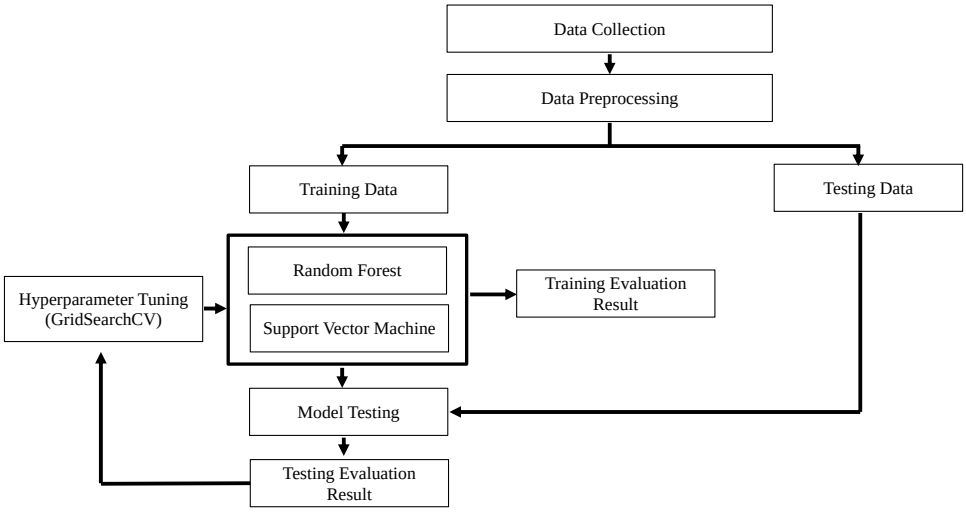


Figure 1. Proposed Model

Data Collection and Data Pre-processing

Weekly sales data were collected from retail SMEs operating in the Surakarta region and surrounding areas of Indonesia between January 2023 and December 2024, yielding approximately 1,000 weekly records. Collected variables included week index, product category, average price, discount/promotion status, holiday/seasonal indicator, prior-week quantity sold (lag feature), and quantity sold (the target variable).

Pre-processing included handling of missing values, one-hot encoding of categorical variables (product category, promotion status), and min-max normalisation of numerical features to a 0–1 range to ensure uniform feature scaling prior to model training. The dataset was then split into training and test subsets using an 80:20 ratio.

Model Development

Two base learners were developed: Random Forest and Support Vector Regression (SVR). Both were optimised through Grid-Search hyperparameter tuning with cross-validation—covering the number of trees and maximum depth for RF, and the kernel type, regularisation parameter (C), and gamma for SVM. Unlike prior hybrid RF-SVM studies that generally relied on default parameter configurations, every hyperparameter combination in this study was systematically searched to identify the configuration that minimised validation error.

Once each base learner was tuned, its predictions were combined through a validation-optimised weighted hybridisation scheme, formalised in Equation (1):

$$\hat{Y}hybrid = \alpha \cdot \hat{Y}RF\text{-}tuned + (1 - \alpha) \cdot \hat{Y}SVM - tuned \quad (1)$$

Where Ŷhybrid denotes the final hybrid prediction, ŶRF-tuned and ŶSVM-tuned denote the predictions of the tuned RF and SVM models respectively, and α is the weighting coefficient determined through iterative search on the validation set to minimise Mean Squared Error (MSE). The key methodological distinction from prior hybrid RF-SVM schemes is that both base learners are optimised prior to hybridisation, rather than

combined at default settings, so that each learner's contribution to the final prediction reflects its best achievable performance.

Model Evaluation

Model performance was assessed using three standard regression metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). MSE and RMSE quantify prediction error magnitude, with RMSE penalising larger errors more heavily, while R^2 indicates the proportion of variance in actual demand explained by the model.

RESULT AND DISCUSSION

The tuned ensemble RF-SVM model was evaluated on a test set of approximately 200 weekly samples. Results are summarised in Table 1.

Table 1. Evaluation Results of the Tuned Ensemble RF-SVM Model

Model	MSE	RMSE	R^2
Tuned Ensemble RF-SVM (GridSearch)	0.001109	0.033308	0.981786

As shown in Table 1, the tuned ensemble model achieved an MSE of 0.001109 and an RMSE of 0.033308 on normalised data (0–1 scale), indicating a very small average prediction error relative to actual values. An R^2 of 0.981786 indicates that the model explains approximately 98.18% of the variance in weekly demand on the test set, reflecting a high degree of fit to the underlying demand pattern. Figure 2 presents the predicted versus actual quantity sold across the full test set.

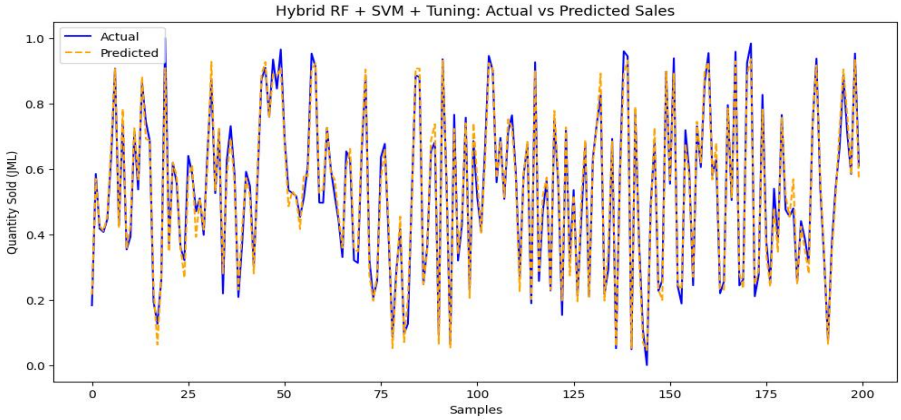


Figure 2. Predicted versus Actual Values of the Tuned Ensemble RF-SVM Model

As shown in Figure 2, the predicted series (orange, dashed) closely tracks the actual series (blue), including at peaks and troughs that reflect seasonal or promotion-driven demand surges. Minor deviations occur at points of sharp fluctuation, but overall the model successfully captures the underlying demand pattern, demonstrating its capacity to handle the non-linear and seasonal characteristics typical of fashion-retail SME demand.

To further characterise the distribution of prediction error, a residual analysis was conducted, shown in Figure 3.

The residual plot in Figure 3 shows that most residuals are randomly scattered around zero without any discernible systematic pattern, indicating no significant bias across the predicted demand range. Residuals fall within approximately -0.06 to 0.08, a narrow

range given the normalised data scale. This random dispersion supports the conclusion that the tuned RF and SVM components complement one another in handling noise and outliers within the weekly demand data.

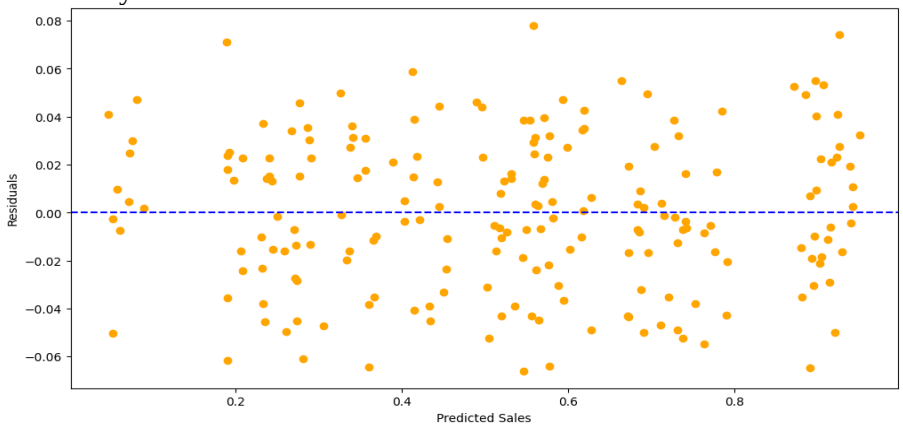


Figure 3. Residual Plot of the Tuned Ensemble RF-SVM Model

To contextualise these findings, Table 2 compares the results of this study with related work using RF- and SVM-based models under default parameter settings. This comparison is necessarily contextual, as each study uses a different dataset and prediction target.

Table 2. Comparison with Related Studies

Study	Model	Reported R^2	Domain
(Ganguly & Mukherjee, 2024)	Random Forest (default)	0.945	General retail
(AbdElminaam et al., 2024)	Random Forest (default)	0.970	Store sales
(Zhu, 2025)	VMD-SVM (default)	0.9214	Retail time-series
This study (2026)	Ensemble RF-SVM + Grid Search Tuning	0.9818	SME weekly demand

From a practical standpoint, these findings suggest that fretail SMEs can use a tuned ensemble RF-SVM model to plan inventory procurement more precisely, reduce the risk of overstocking or stockouts, and align promotional strategy with projected weekly demand. The model's ability to capture seasonal patterns is particularly relevant for operational planning ahead of high-demand periods.

Nonetheless, this study has several limitations. First, Grid-Search hyperparameter tuning is computationally more expensive than using default parameters, which warrants consideration of computational efficiency in large-scale deployments. Second, the hybridisation weight (α) is derived from a specific historical dataset and may require periodic re-tuning as market demand patterns evolve. Third, this study did not benchmark the tuned ensemble RF-SVM model against boosting-based ensembles such as XGBoost, nor against deep learning approaches such as Long Short-Term Memory (LSTM) networks, both of which represent promising directions for future research.

CONCLUSION

This study developed an ensemble RF-SVM model with systematic hyperparameter tuning to predict weekly demand for retail SMEs, and evaluated the extent to which such tuning improves model performance relative to default-parameter baselines reported in the literature. The tuned model achieved an MSE of 0.001109, an RMSE of 0.033308, and an R^2 of 0.981786, indicating that it explains the large majority of variance

in actual demand with small prediction error. Predicted-versus-actual and residual plots further confirm a high degree of model fit and an absence of systematic bias.

The principal contribution of this study lies in explicitly quantifying the value of systematic hyperparameter tuning within an RF-SVM hybridisation scheme—rather than assuming default parameters are sufficient—and in applying this approach to weekly demand-volume forecasting for retail SMEs, a context distinct from prior transaction-value-focused hybrid RF-SVM research. Future work should: (1) benchmark the tuned ensemble against boosting-based ensembles such as XGBoost and deep learning approaches such as LSTM or Temporal Fusion Transformers; (2) explore more computationally efficient tuning strategies such as Bayesian optimisation; (3) extend the dataset to additional SME sectors to test model generalisability; and (4) develop an application interface based on this model to support day-to-day operational decision-making for SME practitioners.

REFERENCES

AbdElminaam, D. S., Mohamed, M. A., Khaled, S., Hany, F., Magdy, M., & Sherif, Y. (2024). Leveraging Machine Learning for Accurate Store Sales Prediction: A Comparative Study. *2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 355–362. https://doi.org/10.1109/miucc62295.2024.10783509

Bhavsar, V., Devi, H., & Dhiman, T. (2025). *Digital Technology Adoption in MSMEs: Challenges, Drivers and Impact on Business Performance*. https://doi.org/10.5281/zenodo.16263400

Chabalala, K., Boyana, S., Kolisi, L., Thango, B. A., & Matshaka, L. (2024). *Digital Technologies and Channels for Competitive Advantage in SMEs: A Systematic Review*. https://doi.org/10.20944/preprints202410.0020.v1

Chen, Z., Liu, J., & Chen, J. (2025). *Machine Learning Methods for Financial Forecasting in Enterprise Planning: Transitioning from Rule-Based Models to Predictive Analytics*. https://doi.org/10.71465/fair527

Ganguly, P., & Mukherjee, I. (2024). Enhancing Retail Sales Forecasting with Optimized Machine Learning Models. *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, 884–889. https://doi.org/10.48550/arxiv.2410.13773

Lai, Y. (2024). *Application and Effectiveness Evaluation of Bayesian Optimization Algorithm in Hyperparameter Tuning of Machine Learning Models*. 351–355. https://doi.org/10.1109/peeec63877.2024.00070

M.V.N, N. (2025). Sales forecasting and demand prediction through time series analysis and machine learning. *International Journal of Apllied Mathematics*, *38*(6s), 1262–1282. https://doi.org/10.12732/ijam.v38i6s.656

Warburton, D. M. (2025). *Enhancing time series forecasting accuracy using hybrid machine learning architectures*. https://doi.org/10.5281/zenodo.16784010

Yuvaraj, V. (2025). An Empirical Study on the Impact of Hyperparameter Tuning on Model Performance. *International Journal For Science Technology And Engineering*, *13*(10), 1084–1089. https://doi.org/10.22214/ijraset.2025.74685

Zhu, A. (2025). VMD-SVM-based Retail Product Demand Forecasting Model. *Scientific Journal of Economics and Management Research.*, *7*(2), 24–31. https://doi.org/10.54691/g7pfr887

Author's short biography

Authors Name:

Name : Eko Purwanto

Affiliation : Universitas Duta Bangsa

Phone Number : 081802525150

Email : eko_purwanto@udb.ac.id

Eko Purwanto, S.Kom., M. Kom., Ph.D, graduated with a bachelor's degree in the Information Systems study program at the Duta Bangsa School of Management and Computer Science (STMIK) Surakarta in 2008. He completed his Master's Degree in the Informatics Engineering Master's Study Program at the School of Information and Computer Management (AMIKOM) Yogyakarta in 2013. He is currently a permanent lecturer and the head of the Undergraduate Study Program - Information Systems at the Faculty of Computer Science, Duta Bangsa University, Surakarta. He teaches courses in Computer Network Design and Management, Executive Information Systems, IS/IT Strategic Planning, and Database Programming. He has authored several books, including Pascal Programming Language, Visual Programming, and Database Programming. Additionally, he has published numerous research studies in scientific journals, both national and international, in the field of Information Systems. He has successfully completed his doctoral studies (Ph.D.) at a university in Kuala Lumpur, Malaysia.

Authors Name:

Name : Bangun Prajadi Cipto Utomo

Affiliation : Universitas Duta Bangsa

Phone Number : +62 812-2596-1977

Email : bangun_prajadi@udb.ac.id

He is Senior Lecturer at Universitas Duta Bangsa Surakarta, Indonesia. He is graduated with a bachelor's degree in the Biology study program at the UNSOED Purwokerto in 1990. He completed his Master's Degree in the Management Master's Study Program at the STIE AUB Surakarta in 2008. He is Lecturer at the Faculty of Law and Business, Duta Bangsa University, Surakarta. He teaches courses in Management, Marketing, and Entrepreneurship. He has authored several books, including Marketing, Management, and Entreprener. Additionally, he has published numerous research studies in scientific journals, both national and international, in the field of Management. He has successfully completed his doctoral studies (Ph.D.) at a university in Kuala Lumpur, Malaysia

Authors Name:

Name : Hanifah Permatasari

Affiliation : Universitas Duta Bangsa

Phone Number : +62 853-3669-9993

Email : hanifah_permatasari@udb.ac.id

She is Senior Lecturer at Universitas Duta Bangsa Surakarta, Indonesia. She is graduated with a bachelor's degree in the Information Systems study program at the Duta Bangsa School of Management and Computer Science (STMIK) Surakarta in 2014. She completed his Master's Degree in the Informatics Engineering Master's Study Program at the UII Yogyakarta in 2019. She is currently a permanent lecturer and the Secretary of the Undergraduate Study

Program - Information Systems at the Faculty of Computer Science, Duta Bangsa University, Surakarta. She teaches courses in Management Information System, E-Business, Database Programming, and Web Programming. She has authored several books, including Management Information System, Database, and E-Business. Additionally, She has published numerous research studies in scientific journals, both national and international, in the field of Information Systems.

Authors Name:

Name : Farahwahida Mohd

Affiliation : Universiti Kuala Lumpur

Phone Number : +60 19-256 2167

Email : farahwahidam@unikl.edu.my

She is Senior Lecturer at Universiti Kuala Lumpur, Malaysia