A Machine Learning-Based Approach for Retail Demand Forecasting: The Impact of Spending Score and Algorithm Optimization

Ni Putu Erica Puspita Adriani[1], Roy Rudolf Huizen[2], Dadang Hermawan[2]
[1]Magister Program, Department of Magister of Information Systems, Institute of Teknologi dan Bisnis STIKOM, Bali, Indonesia
[2]Department of Magister of Information Systems, Institute of Teknologi dan Bisnis STIKOM, Bali, Indonesia

ARTICLE INFO

Article history:

Received January 25, 2025
Revised March 29, 2025
Accepted April 21, 2025

Keywords:

Machine Learning;
Random Forest;
Decision Tree;
Support Vector Machine;
Spending Score

ABSTRACT

Demand forecasting in the retail industry remains a critical challenge, with inaccurate predictions leading to substantial inventory inefficiencies, financial losses, and reduced customer satisfaction. Traditional forecasting methods, primarily reliant on historical sales data, often lack the capacity to effectively model the complexities of dynamic consumer behavior and rapid market fluctuations. To address this, this study proposes a refined demand forecasting approach through the introduction of the Spending Score, a novel synthetic feature that synthesizes customer purchase frequency and total spending to augment predictive accuracy. We implement and optimize machine learning algorithms, specifically Random Forest, Decision Tree, and Support Vector Machine (SVM), using rigorous hyperparameter tuning techniques to determine the most effective model for retail demand prediction. Utilizing detailed customer transaction data, this research aims to identify key purchasing patterns that significantly influence demand variability. By integrating the Spending Score into our predictive models, we provide a data-driven framework enabling retailers to optimize inventory management, enhance targeted marketing strategies, and minimize operational inefficiencies. Empirical results demonstrate that the inclusion of the Spending Score leads to more stable and accurate demand forecasts, facilitating improved alignment between supply and market demand. While acknowledging potential limitations, including data scalability issues and the risk of feature-induced bias, future research will explore the integration of real-time data streams, advanced deep learning methodologies, and expanded datasets to further improve predictive capabilities and model adaptability in the continuously evolving retail landscape.

Corresponding Author:

Ni Putu Erica Puspita Adriani,Master Program, Department of Magister Information System, STIKOM Institute of Technology and Business, Bali, Indonesia
Email: 222012003@stikom-bali.ac.id

1. INTRODUCTION

The retail sector operates within a highly dynamic and competitive landscape, where precise demand forecasting is paramount for optimizing inventory management, ensuring cost efficiency, and maintaining customer satisfaction [1]. Conventional demand prediction methodologies, predicated primarily on historical sales data, frequently prove inadequate in capturing the complexities of evolving consumer behavior, seasonal fluctuations, and exogenous market forces. Consequently, retailers often encounter suboptimal inventory levels, resulting in financial losses, obsolescence, and lost revenue opportunities [2]. Machine learning has emerged as a promising approach to enhance demand forecasting by identifying complex patterns within vast datasets [3], [4]. Various models, such as Random Forest, Decision Tree, and Support Vector Machine (SVM), have been applied to predict product demand in retail sectors [5]. Inaccuracy in projecting demand can lead to overstocking which increases storage costs, or understocking which results in lost customers and revenue [6].

However, existing studies often overlook the impact of synthetic features that combine multiple customer behavioral attributes into a single predictive variable. Moreover, hyperparameter optimization techniques remain underexplored, which affects the overall performance of predictive models [7], [8].

This study introduces Spending Score, a novel synthetic feature that integrates customer purchase frequency and total spending to improve demand forecasting accuracy. Additionally, this research evaluates and optimizes machine learning models using hyperparameter tuning methods to determine the most effective algorithm for retail demand prediction. The integration of engineered features and model optimization provides a data-driven decision-making framework to support retailers in optimizing inventory levels, marketing strategies, and supply chain efficiency [9]. Previous studies in Table 1, have various limitations even though they provide significant contributions [10].

Table 1. Comparison of Research Trends and Contributions

References	Topics	Methodology	Results	Similarities	Differences
[10]	Predicted impact of weather on retail sales	LASSO, GBM, XGB, SVM, MLP	Weather is significant in improving sales prediction accuracy by 56%.	Both focus on sales analysis	Using weather data, not other factors
[11]	Supplier selection using a hybrid approach	Decision Tree, KNeighbors, Random Forest, Logistic Regression	Supplier selection accuracy improved by considering quality, cost, and risk.	Both use machine learning for decisions	Focus on supplier selection, not other optimizations
[12]	Dynamic inventory management	ARIMA-NARNN, XGBoost	Increase turnover by 1.3% with inventory management optimization method.	Focus on inventory management.	Using combinatorial optimization
[13]	Supplychain optimization	RFM, TURF	Demand forecast accuracy increased by 20%, operational costs decreased.	Both support data-driven decisions	Focus on supply chain, not customer analysis
[14]	Algorithm selection for predictive maintenance	Supervised, unsupervised, semi-supervised, reinforced ML	Provides guidance for non-experts in selecting algorithms for predictive maintenance.	Focus on machine learning based optimization	Prioritize maintenance needs, not predictions
[15]	Prediction of supply chain disruption in the textile industry	SVM, AdaBoost, Naive Bayes, ADASYN	Addressing data imbalance, improving the accuracy of supply chain disruption predictions.	Both support machine learning-based predictions	Focus on textiles, not other common applications
[16]	Multichannel fashion demand forecast	K-means Clustering, ELM, SVR	The KM-ELM model is more accurate than the others, with a MAPE of 0.13% and an RMSE of 0.14.	Both support inventory efficiency	Clustering based, not single analysis
[17]	Retail customer behavior prediction	Logistic Regression, Neural Networks	AUC Neural Networks 72.6%, showing high flexibility for non-linear patterns.	Focus on customer behavior	Based on churn classification, not general analysis
[18]	Retail sales prediction with a combination of models	XGBoost, LightGBM, LSTM, Fusion models	The most accurate fusion model (MAPE 0.046, RMSPE 0.060), integrates tree and LSTM models.	Focus on data-driven prediction optimization	Using model fusion instead of a single algorithm

The research [11]focuses on supplier selection without considering other optimizations. Research [12]used a combinatorial optimization approach to inventory management but did not target customer analysis. The study [13] focuses on the supply chain without delving into customer behavior patterns. Research [14] only provides algorithmic guidance for predictive maintenance, not retail predictive analysis. The study [15] limited to the textile sector, without general application. The research [16] rely on clustering that is limited to inventory analysis. The study [17] focused on customer churn classification, without extensive predictive analytics coverage. Finally, [18] used a fusion model for sales prediction but did not explore new synthetic variables.

This study makes a novel contribution by introducing a synthetic feature called Spending Score, which is designed to integrate the dimensions of customer purchase frequency and amount. Spending Score offers a comprehensive approach to capture customer spending patterns that are not detected by standard variables. This feature combines demographic variables (age, gender), geographic (location), and product attributes (category, seasonality, and customer reviews) into a more informative representation [19].

Spending Score is also optimized through a normalization process to ensure scale consistency in the analysis. In addition to feature development, this study implements model optimization techniques using hyperparameter tuning with GridSearchCV and RandomizedSearchCV. This approach guarantees that algorithms such as Random Forest, Decision Tree, and SVM achieve optimal performance by employing parameter tuning to identify the most suitable hyperparameters for the specific dataset [20], [21], [22], [23]. This process is crucial for selecting the most effective model for predicting future product demand. This comparison is carried out to evaluate model performance based on MSE, R^2 and Cross Validation thus it can provide more accurate and relevant business decisions to support strategic decision making in the retail industry.

2. METHODS

This research employed machine learning techniques to forecast product demand within the retail sector. The methodology encompassed several key stages: data collection, preprocessing, feature engineering, model training, optimization, and evaluation. Each step is designed to ensure valid, relevant, and reliable results. The method used in this study is designed to build an accurate future product demand prediction model by utilizing the synthetic Spending Score feature. This study involves a series of systematic steps, from data processing to model evaluation. Fig. 1 is the flowchart of flow on this study.

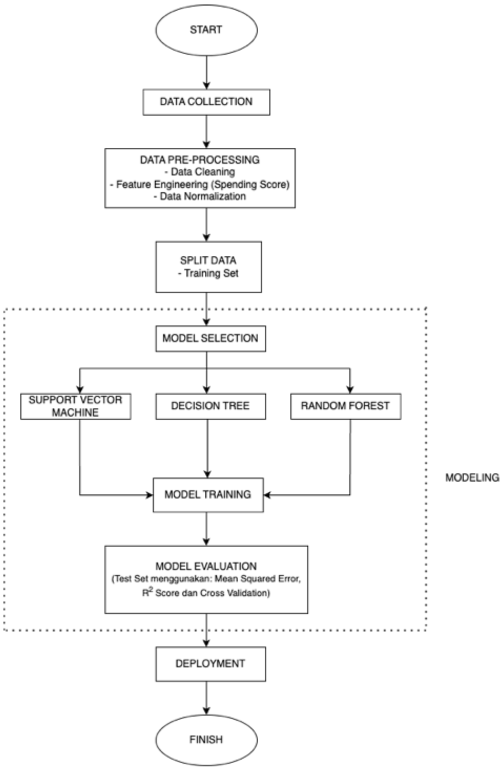


Fig. 1. Flowchart of Methodology Overview

This study in Fig. 1 follows a structured methodology to develop an optimized machine learning-based demand forecasting model for the retail sector. The process begins with data collection, where customer transaction records are gathered. The data undergoes pre-processing, including cleaning, feature engineering, and normalization, to improve quality and predictive accuracy. A key innovation in this step is the introduction of Spending Score, a synthetic feature integrating customer purchase frequency and total spending to enhance demand prediction.The dataset is then split into training and testing subsets to ensure robust evaluation. Next, three machine learning models—Support Vector Machine (SVM), Decision Tree, and Random Forest—are selected based on their ability to capture demand patterns. To optimize model performance, hyperparameter tuning is applied, using GridSearchCV for Decision Tree and SVM and RandomizedSearchCV for Random Forest.

Once trained, the models are evaluated using key metrics such as Mean Squared Error (MSE), R^2 Score and Cross Validation to assess predictive accuracy and classification reliability. The selection of Mean Squared Error (MSE) and R^2 Score as evaluation metrics in this study is based on their ability to provide a comprehensive assessment of model performance in demand forecasting [24], [25], [26]. MSE was chosen over Mean Absolute Error (MAE) due to its sensitivity to large deviations, as it penalizes larger errors more significantly, making it more effective in scenarios where minimizing substantial prediction inaccuracies is crucial. Additionally, MSE was preferred over Root Mean Squared Error (RMSE) because, while RMSE maintains interpretability by preserving the unit of the target variable, it can overly emphasize outliers, which may not always be desirable in retail demand prediction.

Complementing MSE, R^2 Score serves as a measure of how well the model explains variance in the dependent variable, offering a relative performance indicator that facilitates model comparison. A higher R^2 value suggests that the model captures a greater proportion of demand variability, making it a valuable metric alongside MSE [27]. The combination of MSE and R^2 Score ensures both precise error quantification and robust interpretability, making them the most suitable evaluation metrics for optimizing machine learning models in retail demand forecasting.

The best-performing model is then deployed to support retail inventory management, demand forecasting, and marketing strategy optimization. By incorporating Spending Score and hyperparameter tuning, this research enhances predictive accuracy and decision-making efficiency in the retail industry, providing valuable insights for both researchers and practitioners.

2.1. Data Collection and Preparation

This study uses a quantitative approach with detailed datasets in Table 3, consisting of 3,900 entries from the Kaggle site, covering customer demographics, geographic location, and product attributes [19].

Table 3. Dataset Description

Variables	Description	Data Types
Customer ID	Identify each customer as unique.	Integer
Age	Age of customer.	Integer
Gender	Customer gender (male or female).	Object
Item Purchased	Items purchased by customers.	Object
Category	Category of items purchased (e.g. Clothing, Footwear).	Object
Purchased Amount (USD)	The amount of money spent on purchases in USD.	Integer
Location	The geographic location where the customer resides or purchases.	Object
Size	The size of the item purchased (e.g., S, M, L).	Object
Color	The color of the purchase item.	Object
Season	The season in which the purchase was made (e.g., Winter, Spring).	Object
Review Rating	Review rating given by customers for purchased items (scale 1-5).	Float
Subscription Status	Customer subscription status (Yes or No).	Object
Shipping Type	The selected shipping type (e.g., Express, Free Shipping, Next Day Air).	Object
Discount Applied	Whether discount is applied to purchase (Yes or No).	Object
Promo Code Used	Whether the promo code was used on the purchase (Yes or No).	Object
Previous Purchases	The number of previous purchases made by the customer.	Integer
Payment Method	Payment method used (e.g., Venmo, Cash, Credit Card, PayPal).	Object
Frequency of Purchases	Customer purchase frequency (e.g., Fortnightly, Weekly, Annually).	Object

The dataset utilized in this research encompasses a multifaceted array of variables, encompassing diverse data types, to facilitate a comprehensive understanding of customer behavior. The Customer ID variable serves as a unique identifier in the form of a number (Integer) that ensures that there is no repeating customer data. The Age variable, also in the form of a number (Integer), represents the age of the customer and helps in demographic analysis. The Gender variable, in the form of an Object, indicates the gender of the customer (male or female), while the Item Purchased and Category, both in the form of Objects, provide information related to the item and its category, such as clothing or shoes. Furthermore, the Purchased Amount (USD) variable represents the amount of money spent in each transaction in the form of a number (Integer). Location, Size, and Color, all in the form of Objects, provide geographical information, size, and color of the item purchased. Season, also in the form of an Object, records the purchasing season such as winter or spring. Review Rating records customer reviews of the item on a scale of 1–5, using a decimal precision numeric data type (Float).

The Subscription Status, Shipping Type, Discount Applied, and Promo Code Used variables, all in Object form, represent binary information such as subscription status (Yes or No), shipping type, discount applied, and promo code usage. Previous Purchases, with an Integer data type, records the number of previous transactions made by a customer. The Payment Method variable, an Object, records the payment method such as credit card, PayPal, or cash, while Frequency of Purchases, also an Object, describes the frequency of customer transactions (e.g., weekly or yearly). The Transaction Date variable is stored in datetime64 format, allowing for analysis of time-based trends.

2.2. Data Preprocessing

Data pre-processing is crucial for ensuring data quality and preparing the dataset for subsequent analysis. The first step is to clean the data to handle missing values and remove outliers that can distort the analysis [28], [29]. Next, feature engineering is carried out by developing an innovative feature called Spending Score. This feature is designed to provide a deeper picture of consumer behavior by integrating the number of purchases, purchase frequency, and demographic variables. The processed data is divided into training (80%) and testing (20%) subsets for reliable model assessment. To improve the performance of the predictive model by selecting the optimal parameters for each algorithm, hyperparameter tuning is applied using the GridSearchCV and RandomizedSearchCV methods. This process is designed to ensure accurate and relevant results in further analysis.

2.3. Classification Method

The classification methods utilized in this research encompass decision tree, random forest, and support vector machine (SVM) for predicting future product demand. Decision Tree works by building a decision tree based on certain rules that map patterns from input data to predicted results [30], [31], [32]. Random Forest, an ensemble method based on Decision Trees, improves accuracy, and reduces overfitting by combining predictions from multiple trees [33]. Meanwhile, SVM works by finding the best hyperplane that separates data into certain classes, thus it can capture more complex patterns. These three methods were chosen because of their respective advantages in handling heterogeneous data, and their comparison was carried out to determine the best model that can provide the highest prediction accuracy and support more effective business decisions in the retail industry.

2.4. Decision Tree

Decision Tree organizes attributes by ordering them based on their various values. Each decision tree consists of nodes and branches, which are mainly used for classification purposes [34], [35]. Nodes represent attributes in a group, while branches represent the possible values that the node can take. This model utilizes a tree-based framework for decision-making. Each internal node evaluates an input feature, branching to subsequent nodes based on the test result. Ultimately, the model arrives at a leaf node, representing the predicted class or output [36]. The advantages of using this model are that it is easy to understand and interpret because the tree structure and rules generated are very intuitive and easy to understand, it can handle categorical and numeric data, fast training, and prediction time especially for datasets that are not too large, it helps in selecting the most relevant features for prediction [37].

Decision Tree is chosen for its simplicity, interpretability, and efficiency in handling categorical and numerical data. It constructs a hierarchical structure where each node represents a decision rule based on input features, making it an effective model for capturing non-linear relationships in customer purchasing behavior. However, a major limitation of Decision Tree is its tendency to overfit the training data, leading to reduced generalization on unseen data. To mitigate this, pruning techniques and hyperparameter tuning are applied to enhance model robustness [35].

2.5. Random Forest

Ensemble, which improves accuracy and reduces the risk of over-fitting by combining multiple decision trees [34], [38]. The advantages of using Random Forest are that the combination of many decision trees reduces the risk of overfitting compared to a single decision tree, provides more accurate and stable accuracy results compared to a single model, can handle missing values well, produces more stable and consistent predictions through the voting process of several trees, is suitable for datasets that have many features [39].

To overcome the overfitting issues of Decision Tree, Random Forest, an ensemble-based approach, is incorporated. By aggregating multiple decision trees trained on different subsets of data, Random Forest improves prediction stability, reduces variance, and enhances overall accuracy. This method is particularly well-suited for high-dimensional retail datasets where multiple factors such as seasonality, pricing, and customer demographics interact to influence demand. Additionally, Random Forest provides feature importance scores, enabling retailers to identify key variables that drive purchasing trends [40].

2.6. Support Vector Machine

Support Vector Machines are a classification algorithm that maximizes the margin between data classes using an optimal hyperplane [32], [41]. Support Vector Machines (SVMs) exhibit a notable advantage in high-dimensional datasets, effectively handling scenarios where the number of features surpasses the number of samples, can produce a large margin between different classes, can help in the classification of unbalanced data, using kernel functions, SVM can handle nonlinear data well, choosing the right regularization parameters, SVM can avoid overfit, based on strong mathematical theory, so it can be used in various situations.

SVM is selected for its strength in handling high-dimensional data and non-linearly separable patterns. By mapping input features into a higher-dimensional space using kernel functions, SVM efficiently identifies complex demand patterns that may not be captured by tree-based models. This is particularly useful in retail forecasting, where demand fluctuations are influenced by multiple interdependent variables. However, SVM's computational cost increases significantly with large datasets, making it less scalable compared to Random Forest. To address this, hyperparameter tuning and kernel selection are employed to optimize performance while balancing computational efficiency [42], [43].

2.7. Model Training

Following model selection, the chosen model is trained using the designated training set. This training process involves hyperparameter tuning to optimize model performance by iteratively adjusting hyperparameters and evaluating their impact on model behavior [44]. Hyperparameter tuning is crucial for optimizing model performance, balancing accuracy, generalization, and computational efficiency. In this study, RandomizedSearchCV is used for Random Forest due to its high-dimensional hyperparameter space, which includes the number of trees, depth, and feature selection strategy.

Performing exhaustive search with GridSearchCV on such a complex model would be computationally expensive, making stochastic sampling a more efficient approach. In contrast, GridSearchCV is applied to Decision Tree and SVM, as their hyperparameter spaces are more structured and manageable. Decision Tree optimization focuses on tree depth, split criteria, and minimum sample splits, while SVM tuning involves kernel selection, regularization (C), and gamma values. Fine-tuning these parameters enhances model robustness, preventing overfitting in tree-based models and improving decision boundaries in SVM. This strategic approach ensures optimal predictive performance while maintaining computational efficiency, making it well-suited for demand forecasting in retail applications.

2.8. Model Evaluation

Evaluation metrics are employed to assess model performance, specifically its ability to accurately predict outcomes for new, unseen data points. In this study, model evaluation encompasses the rigorous assessment of the machine learning model's performance to ensure its accuracy and reliability in making predictions. The evaluation measures the extent to which the model can generalize patterns from data it has never seen before, using a test dataset that is separate from the training dataset.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

Formula (1), Mean Squared Error (MSE), which quantifies the average squared difference between actual and predicted values, serves as the primary evaluation metric. MSE is employed to assess the predictive accuracy of the models under consideration, including Random Forest, Decision Tree, and Support Vector Machine (SVM), by evaluating their ability to minimize prediction errors on the given dataset [45], [46].

Within this study, MSE serves as a crucial metric for comparing the predictive accuracy of different models. Models with lower MSE values exhibit greater predictive accuracy, demonstrating a stronger alignment between predicted and actual purchase amounts. For example, the Random Forest model has the lowest MSE when compared to the other models, indicating that this model is more accurate and is better at capturing the patterns of the data used in the study.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{2}$$

Formula (2), R^2 or the coefficient of determination (R-squared) quantifies the extent to which the independent variables in the model explain the variability in the dependent variable. Within this research, R-squared serves as a key metric for evaluating model fit, indicating how effectively the model's predictions align with the actual observed values [47], [48].

$$R_{cv}^2 = \frac{1}{k} \sum_{j=1}^{k} R_j^2 \tag{3}$$

Formula (3), or Cross-validation determination coefficient is an evaluation of the performance of the prediction model in a cross-validation scenario. Cross-validation involves calculating the average R-squared value across multiple iterations. This method helps assess the model's performance in differentiating between distinct categories, particularly in identifying specific purchased items [49]. This is relevant to the research objective, which is the prediction of product demand in the retail industry using metrics such as spending score. These metrics support more effective data-driven business decisions by ensuring that the model is not only accurate in providing quantitative predictions of product demand, but also reliable in identifying specific categories.

3. RESULTS AND DISCUSSION

3.1. Results

This study focuses on building a product demand prediction model in the retail sector using machine learning techniques. Demographic, geographic, and product-specific variables (category and season) are used as model inputs. The models tested include Decision Tree, Random Forest, and SVM. Evaluation is carried out using MSE and R^2 metrics to measure model accuracy [50], [51]. The research results provide valuable insights into the factors that influence product demand and can be used to support decision making related to supply chain management and marketing strategies.

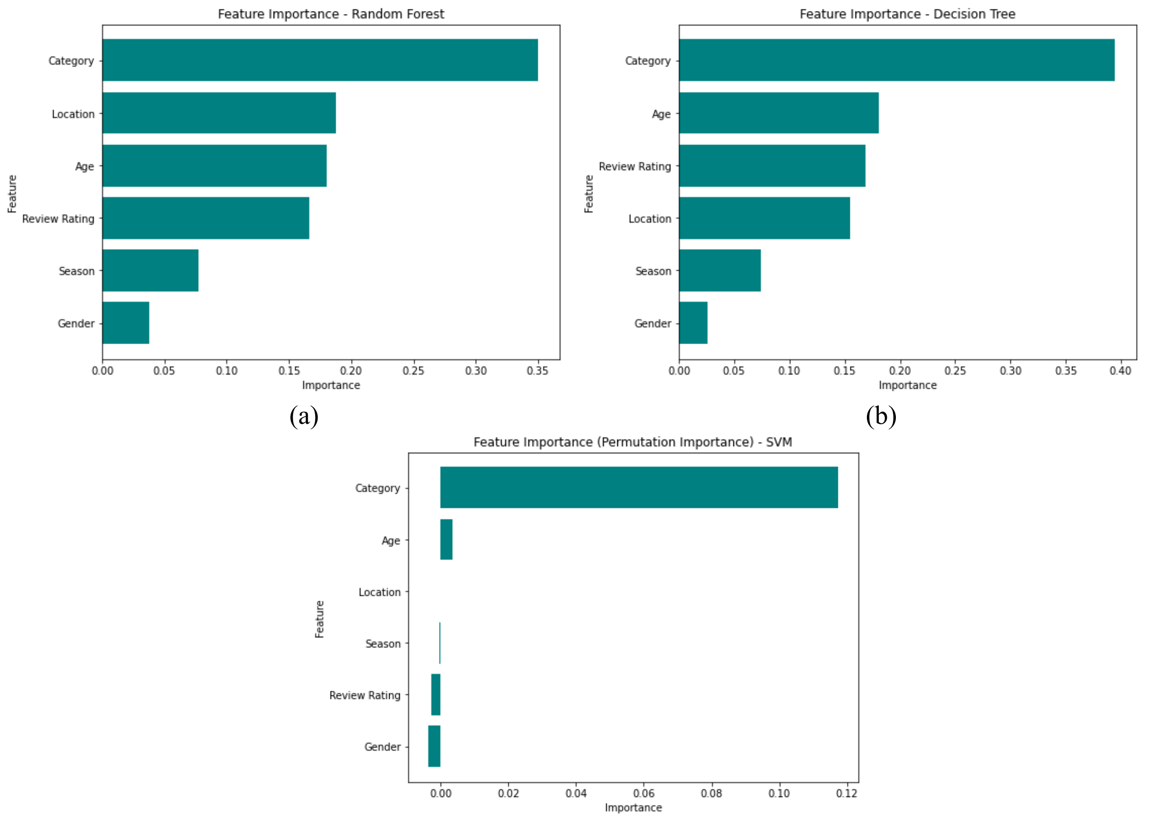
3.2. Spending Score

Spending Score is a new feature in this study designed to represent various factors that influence customer purchasing behavior in the retail industry. This feature is built by combining important variables, such as demographic attributes (age, gender), geographic (location), and product attributes (category, season, and review rating). The process begins with customer data collection, followed by preprocessing such as handling missing values and encoding categorical variables. The main variables are weighted based on their influence on shopping behavior, then used to calculate the Spending Score as a composite score. To enhance model performance, the score was normalized using MinMaxScaler, ensuring consistent scaling across all data points. This normalized score is now suitable for input into machine learning models. Furthermore, the relative influence of each variable is explicitly defined through assigned weights, with the total weight sum equaling 100%.

Spending Score is a new feature designed to combine variables such as age, gender, location, product category, season, and customer reviews that influence shopping behavior. Each variable is weighted based on its level of influence, with product category having the highest weighting of 20%, as it is considered the most influential factor. Meanwhile, variables such as age, gender, and location are each given a weighting of 10%, while season and customer reviews are only given a weighting of 5%, reflecting their lesser influence. These weights are designed to provide a proportional contribution of each variable in the calculation of Spending Score, reflecting the characteristics of the retail domain, and can be adjusted to improve prediction accuracy.

A fixed weighting scheme may lead to model distortions, particularly if it overemphasizes certain purchasing factors while underrepresenting others. This can impact inventory planning and marketing strategies, reducing the model's generalizability across diverse customer segments and product categories. To enhance objectivity, data-driven techniques such as feature importance ranking, correlation analysis, or adaptive weighting through machine learning algorithms should be explored. These methods would allow the model to dynamically adjust Spending Score composition based on empirical evidence rather than fixed assumptions, improving both predictive reliability and cross-market applicability.

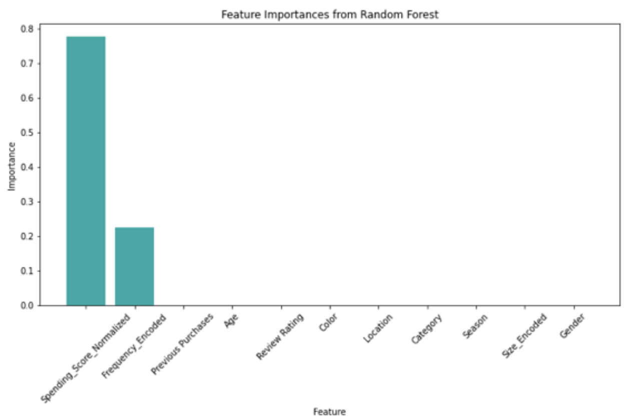
Before implementing the Spending Score, feature importance on Fig. 2 analysis indicated that the 'Category' feature exerted the strongest influence on model predictions for all three algorithms: Random Forest, Decision Tree, and SVM. In the Random Forest and Decision Tree models, in addition to "Category", "Location", "Age", and "Review Rating" features also show significant contributions to predictions, with the "Season" and "Gender" features having a lower influence. In contrast, the SVM model shows the full dominance of the "Category" feature, while other features, such as "Age", "Location", "Season", and "Review Rating", only make very small contributions.



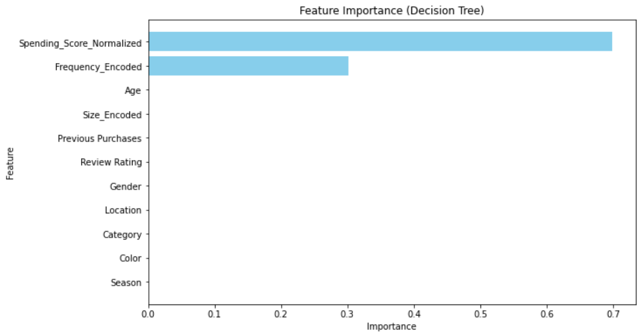
(c)

Fig. 2. Feature Importance Before Using Spending Score (a) Random Forest (b) Decision Tree (c) SVM

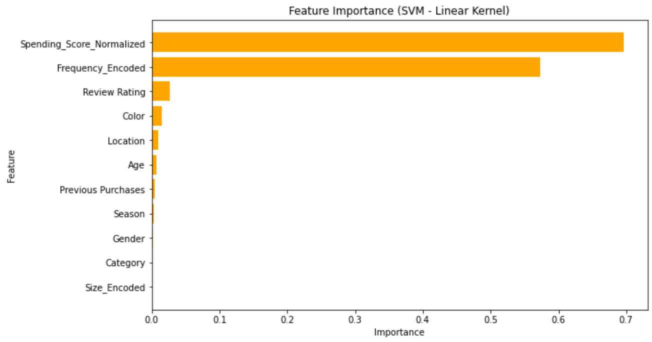
This indicates by Fig. 2 that, without Spending Score, the model relies more on the purchase category as the main factor in prediction. Random Forest and Decision Tree show a more even distribution of feature importance compared to SVM, which significantly utilizes only one dominant feature. This finding suggests the need for additional features such as Spending Score to help the model capture more complex and diverse prediction patterns. The results on Fig. 3 of the Feature Importance analysis show that Spending Score and Frequency Encoded are the main features that have the most influence in the three Machine Learning models, namely SVM, Random Forest, and Decision Tree.



(a)



(b)



(c)

Fig. 3. Feature Importance After Using Spending Score (a) Random Forest (b) Decision Tree (c) SVM

Spending Score dominates the prediction contribution in all models, especially in Random Forest with an influence of almost 80%. Other features, such as Review Rating, Color, and Location, have very little impact, indicating low relevance in the analysis. Overall, customer shopping behavior, especially spending and purchase frequency, are the main indicators in determining the prediction results. Business strategies should focus on the features with the greatest influence to improve prediction accuracy and decision making.

3.3. Optimization Model

Model optimization is the process of improving the performance of a machine learning model by adjusting parameters (both hyperparameters and internal parameters) to produce more accurate predictions. Random Forest uses RandomizedSearchCV which is a random hyperparameter search method to find the best combination based on scoring. Table 4 shows the parameter tuning test of Random Forest. Before optimization was carried out on the three models used in this study, the evaluation results were displayed as in Table 4 attached.

Table 4. Evaluation Results Before Using Optimization

Method	MSE	R^2	Cross Validation
Decision Tree	88.71	-0.72	0.17
Random Forest	76.25	-0.48	0.16
Support Vector Machine	84.73	-0.64	0.17

Table 4 illustrates the evaluation of model performance before optimization using MSE, R^2, and Cross Validation metrics. Random Forest shows the best performance compared to Decision Tree and Support Vector Machine (SVM), with MSE of 76.25 and R^2 -0.48, although the negative R^2 value indicates that the model is not optimal in explaining data variability. Meanwhile, Decision Tree has the highest MSE of 88.71, followed by SVM with MSE of 84.73. Similar Cross Validation values in the three models (around 0.16–0.17) indicate consistent model stability, but are still low. These results are the basis for further optimization to improve model performance.

The discussion of the parameter tuning test range begins with determining the model parameters to be optimized in the training process. Parameter tuning aims to improve model performance by selecting a combination of hyperparameters that provide the best results based on certain evaluations, such as R^2 Score or

Mean Squared Error (MSE). The range of values used in tuning must be chosen carefully to cover the best possible combinations, while still considering computational time efficiency.

Table 5, parameter selection in RandomizedSearchCV for Random Forest is designed to explore model configurations ranging from simple to complex to achieve optimal performance. The n_estimators parameter is selected in the range [50, 100, 200, 300, 400] to balance training time and accuracy, while max_depth is set in [5, 10, 15, 20, None] to manage bias and variance. min_samples_split and min_samples_leaf are adjusted in the range [2, 5, 10] and [1, 2, 4] respectively to control node splitting and prevent overfitting. With RandomizedSearchCV, 20 parameter iterations are tested randomly (n_iter=20), supported by 5-fold cross-validation (cv=5) for reliable evaluation. This approach ensures a model that is accurate, efficient, and has good generalization.

Table 5. Random Forest Parameter Tuning Test Range

Random Forest Tunning Parameters	Test Range
n_estimators	50, 100, 200, 300, 400
max_depth	5, 10, 15, 20, None
min_samples_split	2, 5, 10
min_samples_leaf	1, 2, 4

While effective in demand forecasting, the selected models have inherent limitations. Random Forest is prone to overfitting, especially with a small dataset and the influence of the Spending Score. Its ensemble nature captures complex patterns but risks poor generalization. To mitigate this, regularization techniques, including limiting tree depth and reducing estimators, were applied. Additionally, negative R^2 scores before optimization indicated that the models initially performed worse than a mean-based predictor, likely due to high variance and suboptimal hyperparameters. This issue was addressed through feature normalization and refined hyperparameter tuning, improving model stability.

Table 6, the selection of weights in param_grid aims to explore the influence of hyperparameters on the performance of the Decision Tree model. max_depth is set in [1, 3, 10, None] to balance model simplicity and the ability to capture data patterns. min_samples_split ([2, 5, 7]) and min_samples_leaf ([1, 2, 4]) set the minimum number of samples for node and leaf splits, with small values capturing more details and large values improving generalization. The class_weight parameter ([None, 'balanced']) evaluates the effect of adjusting class weights on data imbalance. This process uses GridSearchCV with 5-fold cross-validation (cv=5) and the f1_weighted metric to ensure reliable evaluation and optimization of the Decision Tree model performance.

Table 6. Decision Tree Parameter Tunning Test Range

Decision Tree Tunning Parameters	Test Range
max_depth	1, 3, 10, None
min_samples_split	2, 5, 7
min_samples_leaf	1, 2, 4
class_weight	None, 'balanced'

Table 7, Parameter selection in param_grid for SVM optimization using GridSearchCV is designed to explore combinations that affect model performance. The C parameter ([0.1, 1, 10]) controls regularization, with small values producing simple models and large values providing higher flexibility. The gamma parameter ([1, 0.1, 0.01]) controls the influence of data points, with large values focusing on local patterns, while small values avoid overfitting with a global approach. Kernels are tested with 'rbf' for non-linear patterns and 'linear' for linear patterns. Three-fold cross-validation (cv=3) and the f1_weighted metric is used to ensure fair evaluation on imbalanced datasets, thus finding the optimal combination to maximize SVM performance.

The poor performance of SVM likely stems from kernel selection issues, data distribution challenges, and computational complexity. If the chosen kernel failed to capture non-linear relationships, the model's ability to generalize would be compromised. Additionally, class imbalance may have further reduced SVM's effectiveness in predicting minority demand categories. Alternative feature transformations or ensemble techniques could improve its performance.

Table 7. Support Vector Machine Parameter Tunning Test Range

SVM Tuning Parameters	Test Range
C	0.1, 1, 10
gamma	1, 0.1, 0.01
kernel	'rbf', 'linear'

3.4. Discussion

After getting the results of this research is the process of training machine learning algorithms, such as Random Forest, Decision Tree, and Support Vector Machine (SVM), to learn patterns from input data (features) and output (targets). This process uses processed training data, where the model tries to minimize prediction errors by optimizing its internal parameters.

Table 8, model evaluation using the Mean Squared Error (MSE), R^2 Score, and Cross-Validation R^2 Score metrics shows that Random Forest provides the best performance with MSE 6.13, R^2 Score 0.93, and Cross-Validation R^2 Score 0.94. This indicates its excellent ability to explain data variability and produce consistent predictions. Decision Tree is in second place with MSE 5.47 and R^2 Score 0.91, showing quite good performance, although not as good as Random Forest. SVM has the lowest performance with MSE 18.95 and R^2 Score 0.67, indicating that the model is less than optimal in capturing data patterns. Overall, Random Forest is the most accurate and reliable model in this evaluation.

Table 8. Evaluation Results After Using Optimization

Method	MSE	R^2	Cross Validation
Decision Tree	5.47	0.91	0.91
Random Forest	6.13	0.93	0.94
Support Vector Machine	18.95	0.67	0.61

Regarding evaluation metrics, while MSE and R^2 Score offer insights into prediction accuracy, MSE over-penalizes large errors, and R^2 alone does not capture absolute error magnitude. Alternative metrics like MAE or RMSE could provide a more balanced assessment, as MAE is more interpretable, and RMSE retains unit consistency, leading to a more comprehensive evaluation of forecasting performance.

Fig. 4, Comparison of results after optimization shows Random Forest with the best performance compared to Decision Tree and SVM, with the lowest MSE, highest R^2 Score, and stable Cross Validation R^2, indicating good accuracy and generalization. Decision Tree is in second place with lower MSE and R^2 Score, but still better than SVM. SVM has the worst performance with the highest MSE and lowest R^2 Score, indicating its inability to capture data patterns. Therefore, Random Forest is recommended as the best model for this dataset.

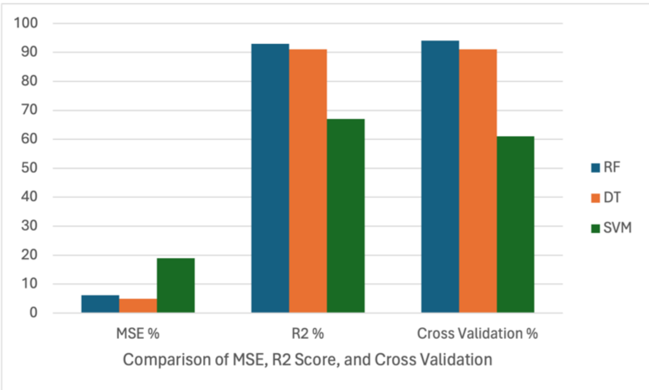
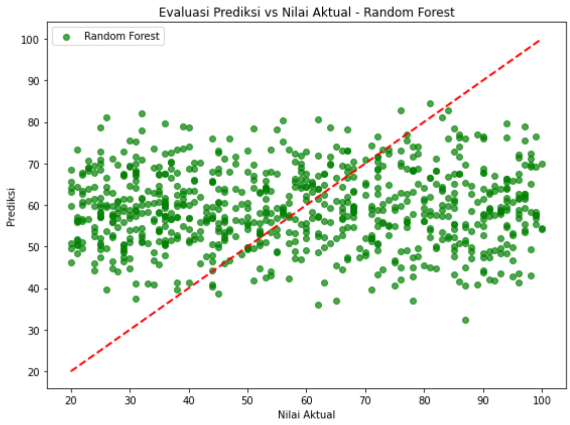


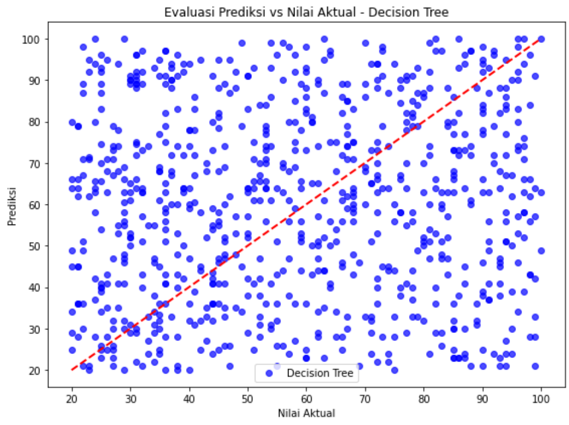
Fig. 4. Comparison of Results After Optimize

Based on the visualization results in Fig. 5, the Random Forest model shows the best performance with predictions closest to the diagonal line, which is the perfect prediction line. Most of the prediction points in this model are around the line, indicating high accuracy and the model's ability to capture data patterns consistently. On the other hand, the Decision Tree model has a prediction spread that is further from the diagonal line, indicating a higher level of prediction error than Random Forest, although it is still quite good at explaining data variability.

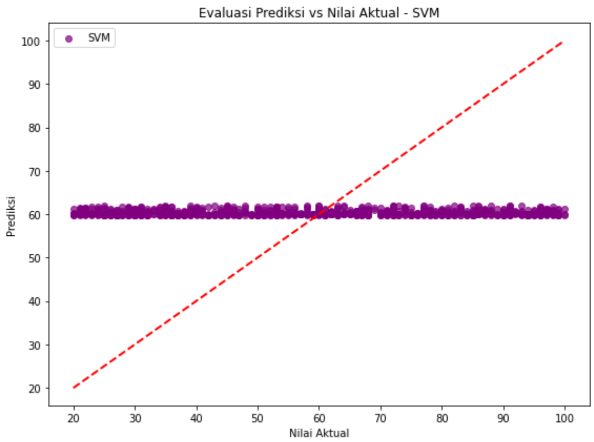
Meanwhile, the Support Vector Machine (SVM) model showed the worst performance, with predictions that were almost constant around a certain value, so far from the perfect prediction line. This shows that SVM failed to capture the data pattern well. Overall, Random Forest can be recommended as the best model for this dataset, followed by Decision Tree, while SVM is less than optimal for use in this context.



(a)



(b)



(c)

Fig. 5. Evaluation of Prediction vs Actual Values of Model (a) Random Forest (b) Decision Tree (c) SVM

4. CONCLUSION

This study highlights the effectiveness of machine learning models, particularly Random Forest, Decision Tree, and SVM, in improving retail demand forecasting through feature engineering and hyperparameter optimization. The introduction of Spending Score as a synthetic feature demonstrates its potential to enhance predictive accuracy; however, several limitations must be considered. Overfitting remains a concern, especially in Random Forest, due to the relatively small dataset, which may reduce the model's generalization ability on larger and more diverse data. Additionally, the validity of the Spending Score weighting scheme requires further evaluation, as its current formulation, while effective in this study, may introduce bias when applied to different retail environments. Furthermore, the SVM model underperformed, likely due to kernel selection issues and the non-linearity of customer purchasing behaviors, suggesting the need for alternative kernel functions or ensemble-based enhancements to improve its predictive performance.

While the results indicate that the proposed approach enhances demand forecasting, it is essential to avoid overgeneralization without extensive validation. The model cannot be assumed to be universally applicable across all retail sectors without further testing on varied datasets. Additionally, its scalability must be assessed, particularly regarding computational feasibility and predictive stability on larger datasets. Future research should explore whether the model remains effective under different market conditions, customer behaviors, and product categories.

To further refine the approach, ensemble models or deep learning architectures should be investigated to enhance predictive accuracy. Additionally, Spending Score weighting should be reevaluated using data-driven statistical methods to improve adaptability across various datasets. Finally, testing the model on larger and more diverse datasets is necessary to evaluate its generalization capability and real-world applicability. Addressing these limitations will contribute to the ongoing advancement of data-driven demand forecasting in the retail sector.

Acknowledgements

The author gratefully acknowledges the Institute of Technology and Business STIKOM Bali for providing the necessary research facilities that facilitated the completion of this study. Deep appreciation is extended to all those who generously contributed relevant datasets and technical resources, enabling the development and evaluation of the machine learning models. This invaluable support has significantly contributed to the in-depth exploration of data management and decision-making within the retail sector through the application of machine learning techniques.

REFERENCES

[1] L. L. Har, U. K. Rashid, L. Te Chuan, S. C. Sen, and L. Y. Xia, "Revolution of Retail Industry: From Perspective of Retail 1.0 to 4.0," in *Procedia Computer Science*, pp. 1615–1625, 2022, https://doi.org/10.1016/j.procs.2022.01.362.

[2] M. S. Sousa, A. L. D. Loureiro, and V. L. Miguéis, "Predicting demand for new products in fashion retailing using censored data," *Expert Syst Appl*, vol. 259, Jan. 2025, https://doi.org/10.1016/j.eswa.2024.125313.

[3] R. M. van Steenbergen and M. R. K. Mes, "Forecasting demand profiles of new products," *Decis Support Syst*, vol. 139, Dec. 2020, https://doi.org/10.1016/j.dss.2020.113401.

[4] K. Javaid *et al.*, "Explainable Artificial Intelligence Solution for Online Retail," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 4425–4442, 2022, https://doi.org/10.32604/cmc.2022.022984.

[5] A. A. Khade, "Performing Customer Behavior Analysis using Big Data Analytics," in *Procedia Computer Science*, Elsevier B.V., 2016, pp. 986–992. https://doi.org/10.1016/j.procs.2016.03.125.

[6] H. S. Le *et al.*, "Predictive model for customer satisfaction analytics in E-commerce sector using machine learning and deep learning," *International Journal of Information Management Data Insights*, vol. 4, no. 2, Nov. 2024, https://doi.org/10.1016/j.jjimei.2024.100295.

[7] A. Munde and J. Kaur, "Predictive Modelling of Customer Sustainable Jewelry Purchases Using Machine Learning Algorithms," in *Procedia Computer Science*, pp. 683–700, 2024, https://doi.org/10.1016/j.procs.2024.04.066.

[8] T. Falatouri, F. Darbanian, P. Brandtner, and C. Udokwu, "Predictive Analytics for Demand Forecasting - A Comparison of SARIMA and LSTM in Retail SCM," in *Procedia Computer Science*, pp. 993–1003, 2022, https://doi.org/10.1016/j.procs.2022.01.298.

[9] A. ul Husna, S. H. Amin, and A. Ghasempoor, "Machine Learning Techniques and Multi-objective Programming to Select the Best Suppliers and Determine the Orders," *Machine Learning with Applications*, p. 100623, Jan. 2025, https://doi.org/10.1016/j.mlwa.2025.100623.

[10] H. Chan and M. I. M. Wahab, "A machine learning framework for predicting weather impact on retail sales," *Supply Chain Analytics*, vol. 5, p. 100058, Mar. 2024, https://doi.org/10.1016/j.sca.2024.100058.

[11] A. Abdulla and G. Baryannis, "A hybrid multi-criteria decision-making and machine learning approach for explainable supplier selection," *Supply Chain Analytics*, vol. 7, p. 100074, Sep. 2024, https://doi/org/10.1016/j.sca.2024.100074.

[12] K. Namir, H. Labriji, and E. H. Ben Lahmar, "Decision Support Tool for Dynamic Inventory Management using Machine Learning, Time Series and Combinatorial Optimization," in *Procedia Computer Science*, pp. 423–428, 2021, https://doi.org/10.1016/j.procs.2021.12.264.

[13] R. Zimmermann and P. Brandtner, "From Data to Decisions: Optimizing Supply Chain Management with Machine Learning-Infused Dashboards," in *Procedia Computer Science*, pp. 955–964, 2024, https://doi.org.10.1016/j.procs.2024.05.184.

[14] S. Arena, E. Florian, F. Sgarbossa, E. Sølvsberg, and I. Zennaro, "A conceptual framework for machine learning algorithm selection for predictive maintenance," *Eng Appl Artif Intell*, vol. 133, Jul. 2024, https://doi.org/10.1016/j.engappai.2024.108340.

[15] I. Jebbor, Z. Benmamoun, and H. Hachimi, "Forecasting supply chain disruptions in the textile industry using machine learning: A case study," *Ain Shams Engineering Journal*, vol. 15, no. 12, pp. 103116, Dec. 2024, https://doi.org/10.1016/j.asej.2024.103116.

[16] I. F. Chen and C. J. Lu, "Demand forecasting for multichannel fashion retailers by integrating clustering and machine learning algorithms," *Processes*, vol. 9, no. 9, Sep. 2021, https://doi.org/10.3390/pr9091578.

[17] S. Akhmetbek, "Forecasting Customer Future Behavior In Retail Business Using Machine Learning Models," *Scientific Journal of Astana IT University*, vol. 10, pp. 45–60, Jun. 2022, https://doi.org/10.37943/ILMM7870.

[18] Y. Sun, "Sales Prediction Based on Machine Learning Approach," *International Conference on Social Sciences and Economic Development (ICSSED),* pp. 1016–1023, 2024, https://doi.org/10.2991/978-94-6463-459-4_113.

[19] Z. Solver, "Consumer Behavior and Shopping Habits Dataset," Kaggle, 2021. [Online]. https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset.

[20] Ç. Demirel, A. A. Tokuç, and A. T. Tekin, "Click prediction boosting via Bayesian hyperparameter optimization-based ensemble learning pipelines," *Intelligent Systems with Applications*, vol. 17, Feb. 2023, https://doi.org/10.1016/j.iswa.2023.200185.

[21] H. Lu, X. Chen, C. Fang, and H. Yang, "Data analytics for optimizing extreme fast charging: a survey," *Data Science and Management*, vol. 1, no. 1, pp. 23–31, Mar. 2021, https://doi.org/10.1016/j.dsm.2021.02.001.

[22] X. Xu and Y. Zhang, "A Gaussian process regression machine learning model for forecasting retail property prices with Bayesian optimizations and cross-validation," *Decision Analytics Journal*, vol. 8, Sep. 2023, https://doi.org/10.1016/j.dajour.2023.100267.

[23] H. Wijaya, D. P. Hostiadi, and E. Triandini, "Optimization XGBoost Algorithm Using Parameter Tunning in Retail Sales Prediction," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 13, no. 3, Dec. 2024, https://doi.org/10.23887/janapati.v13i3.82214.

[24] S. Chatterjee, R. Chaudhuri, S. Gupta, U. Sivarajah, and S. Bag, "Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm," *Technol Forecast Soc Change*, vol. 196, Nov. 2023, https://doi.org/10.1016/j.techfore.2023.122824.

[25] M. Seyedan and F. Mafakheri, "Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities," *J Big Data*, vol. 7, no. 1, 2020, https://doi.org/10.1186/s40537-020-00329-2.

[26] O. Olubusola, N. Z. Mhlongo, D. O. Daraojimba, A. O. Ajayi-Nifise, and T. Falaiye, "Machine learning in financial forecasting: A U.S. review: Exploring the advancements, challenges, and implications of AI-driven predictions in financial markets," *World Journal of Advanced Research and Reviews*, vol. 21, no. 2, pp. 1969–1984, Feb. 2024, https://doi.org/10.30574/wjarr.2024.21.2.0444.

[27] N. Zougagh, A. Charkaoui, and A. Echchatbi, "Artificial intelligence hybrid models for improving forecasting accuracy," in *Procedia Computer Science*, pp. 817–822, 2021, https://doi.org/10.1016/j.procs.2021.04.013.

[28] A. Hayat, M. U. Ashraf, M. Sajawal, S. Usman, and H. S. Alshaikh, "Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques," *Lahore Garrison University Research Journal of Computer Science and Information Technology*, vol. 6, no. 4, pp. 33-45, 2023, https://doi.org/10.54692/lgurjcsit.2022.06004399.

[29] X. Dairu and Z. Shilong, "Machine Learning Model for Sales Forecasting by Using XGBoost," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE,* pp. 480–483, Jan. 2021, https://doi.org/10.1109/ICCECE51280.2021.9342304.

[30] L. Ji and S. Li, "A Dynamic Financial Risk Prediction System for Enterprises Based on Gradient Boosting Decision Tree Algorithm," *Systems and Soft Computing*, p. 200189, Jan. 2025, https://doi.org/10.1016/j.sasc.2025.200189.

[31] B. K. Meher, M. Singh, R. Birau, and A. Anand, "Forecasting stock prices of fintech companies of India using random forest with high-frequency data," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 1, Mar. 2024, https://doi.org/10.1016/j.joitmc.2023.100180.

[32] M. H. Almaspoor, A. Safaei, A. Salajegheh, B. Minaei-Bidgoli, and A. A. Safaei, "Support Vector Machines in Big Data Classiication: A Systematic Literature Review," 2021, https://doi.org/10.21203/rs.3.rs-663359/v1.

[33] S. K. Kiangala and Z. Wang, "An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment," *Machine Learning with Applications*, vol. 4, p. 100024, Jun. 2021, https://doi.org/10.1016/j.mlwa.2021.100024.

[34] B. Talekar, "A Detailed Review on Decision Tree and Random Forest," *Biosci Biotechnol Res Commun*, vol. 13, no. 14, pp. 245–248, Dec. 2020, https://doi.org/10.21786/bbrc/13.14/57.

[35] S. Ravi Teja Reddy and P. Malathi, "Design and Implementation of Sales Prediction Model Using Decision Tree Regressor over Linear Regression Towards Increase in Accuracy of Prediction," in *Advances in Parallel Computing*, pp. 522–529, 2022, https://doi.org/10.3233/APC220074.

[36] Q. Li, X. Wang, Q. Pei, X. Chen, and K.-Y. Lam, "Consistency preserving database watermarking algorithm for decision trees," *Digital Communications and Networks*, Jan. 2023, https://doi.org/10.1016/j.dcan.2022.12.015.

[37] S. Sitanskiy, L. Sebastia, and E. Onaindia, "Learning and Recognizing Human Behaviour with Relational Decision Trees," *Procedia Comput Sci*, vol. 246, pp. 3168–3177, 2024, https://doi.org/10.1016/j.procs.2024.09.356.

[38] J. Zhang, "Impact of an improved random forest-based financial management model on the effectiveness of corporate sustainability decisions," *Systems and Soft Computing*, vol. 6, Dec. 2024, https://doi.org/10.1016/j.sasc.2024.200102.

[39] X. Ma *et al.*, "Predicting the utilization factor of blasthole in rock roadways by random forest," *Underground Space (China)*, vol. 11, pp. 232–245, Aug. 2023, https://doi.org/10.1016/j.undsp.2023.01.006.

[40] A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, and G. Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique," in *Procedia Computer Science*, pp. 806–813, 2021, https://doi.org/10.1016/j.procs.2022.01.100.

[41] N. H. Putri, M. Fatekurohman, and I. M. Tirta, "Credit risk analysis using support vector machines algorithm," in *Journal of Physics: Conference Series*, vol. 1836, no. 1, p. 012039, Mar. 2021, https://doi.org/10.1088/1742-6596/1836/1/012039.

[42] M. A. Almaiah *et al.*, "Performance Investigation of Principal Component Analysis for Intrusion Detection System Using Different Support Vector Machine Kernels," *Electronics (Switzerland)*, vol. 11, no. 21, Nov. 2022, https://doi.org/10.3390/electronics11213571.

[43] M. M. Dakwah, A. A. Firdaus, F. Furizal, and R. Faresta, "Sentiment Analysis on Marketplace in Indonesia using Support Vector Machine and Naïve Bayes Method," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 1, p. 39, Feb. 2024, https://doi.org/10.26555/jiteki.v10i1.28070.

[44] A. H. Alzraiee and R. G. Niswonger, "A probabilistic approach to training machine learning models using noisy data," *Environmental Modelling and Software*, vol. 179, Aug. 2024, https://doi.org/10.1016/j.envsoft.2024.106133.

[45] N. Sharma, R. Sharma, and N. Jindal, "Machine Learning and Deep Learning Applications-A Vision," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 24–28, Jun. 2021, https://doi.org/0.1016/j.gltp.2021.01.004.

[46] D. Raditya, E. P. Nicholas, A. S. Ferarida, and N. Hanafiah, "Predicting Sneaker Resale Prices using Machine Learning," in *Procedia Computer Science*, pp. 533–540, 2021, https://doi.org/10.1016/j.procs.2021.01.037.

[47] N. Patankar, S. Dixit, A. Bhamare, A. Darpel, and R. Raina, "Customer Segmentation Using Machine Learning," *Advances in Parallel Computing*, vol. 39, pp. 239–244, 2021, https://doi.org/10.3233/APC210200.
[48] P. Ganguly and I. Mukherjee, "Enhancing Retail Sales Forecasting with Optimized Machine Learning Models," *International Conference on Sustainable Expert Systems (ICSES)*, pp. 884-889, Oct. 2024, https://doi.org/10.1109/ICSES63445.2024.10762950.
[49] M. Syberg, N. West, D. Lenze, and J. Deuse, "Framework for predictive sales and demand planning in customer-oriented manufacturing systems using data enrichment and machine learning," in *Procedia CIRP*, pp. 1107–1112, 2023, https://doi.org/10.1016/j.procir.2023.09.133.
[50] Z. N. Jawad and V. Balázs, "Machine learning-driven optimization of enterprise resource planning (ERP) systems: a comprehensive review," *Beni Suef Univ J Basic Appl Sci*, vol. 13, no. 1, 2024, https://doi.org/10.1186/s43088-023-00460-y.
[51] D. Falkner, M. Bögl, A. Gattinger, R. Stainko, J. Zenisek, and M. Affenzeller, "Integrating Machine Learning into Supply Chain Management: Challenges and Opportunities," in *Procedia Computer Science*, pp. 1779–1788, 2024, https://doi.org/10.1016/j.procs.2024.01.176.

BIOGRAPHY OF AUTHORS

Ni Putu Erica Puspita Adriani, received her bachelor's degree in computer systems from the STIKOM Bali Institute of Technology and Business and is currently pursuing a master's degree in Information Systems at the same institution. Interests focus on Data Science, with an emphasis on applying Machine Learning for data-driven decision-making. Email: 222012003@stikom-bali.ac.id. ORCID: https://orcid.org/0009-0006-1110-1968.

Roy Rudolf Huizen, graduated with a Doctor of Computer Science (2018) from Gadjah Mada University (UGM) Yogyakarta, Indonesia. Lecturer and researcher at the Department of Master Information Systems at the STIKOM Bali Institute of Technology and Business, with research interests in the fields of Object Identification, Signal Processing, Cyber Security Forensics and Artificial Intelligence. Email: roy@stikom-bali.ac.id. ORCID: 0000-0002-3671-6030.

Dadang Hermawan, holds a doctorate in Management from Brawijaya University, a Master's degree in Management from Udayana University, and a Bachelor's degree in Accounting from Padjadjaran University. He has been working at ITB STIKOM Bali since 2002, where he currently holds the position of rector and serves as a lecturer. is actively involved in various professional organizations, such as APTIKOM, INDOCEISS, CORIS, APTISI, and FORUM REKTOR BALI, where he holds key positions such as general treasurer, chairman, and deputy general treasurer. Email: dadang@stikom-bali.ac.id. ORCID: https://orcid.org/0000-0002-9935-8090.