PREDICTIVE ANALYTICS IN CUSTOMER BEHAVIOR: UNVEILING ECONOMIC AND GOVERNANCE INSIGHTS THROUGH MACHINE LEARNING

Rattapol Kasemrat *, Tanpat Kraiwanit *, Nithima Yuenyong **

* Faculty of Economics, Rangsit University, Pathum Thani, Thailand
** *Corresponding author,* International College, Pathumthani University, Pathum Thani, Thailand
Contact details: International College, Pathumthani University, 140 Santi Suk Mu 4 Alley, Ban Klang, Mueang Pathum Thani District, Pathum Thani 12000, Thailand

OPEN ACCESS

How to cite this paper: Kasemrat, R., Kraiwanit, T., & Yuenyong, N. (2025). Predictive analytics in customer behavior: Unveiling economic and governance insights through machine learning [Special issue]. *Journal of Governance & Regulation, 14*(1), 318–331.
https://doi.org/10.22495/jgrv14i1siart8

ISSN Print: 2220-9352
ISSN Online: 2306-6784

Received: 08.05.2024
Revised: 16.08.2024; 29.01.2025
Accepted: 14.02.2025

JEL Classification: C53, C55, D12, M31, L81
DOI: 10.22495/jgrv14i1siart8

Abstract

This study addresses the challenge of predicting customer behavior in the Thai e-commerce sector using machine learning (ML) models. The research aims to evaluate the effectiveness of random forest and gradient boosting models, benchmarking them against traditional models such as logistic regression and linear regression. Utilizing a dataset from a leading Thai e-commerce platform, the study identifies key predictors of customer satisfaction and spending power, including purchase frequency, age, and customer referrals. The findings reveal that ML models outperform traditional methods in capturing complex, non-linear relationships within consumer data. These results suggest practical applications in customer segmentation, personalized marketing, and optimized resource allocation (Chen et al., 2012). The study also emphasizes the importance of ethical considerations, such as transparency and data privacy, in predictive analytics. The research contributes to the integration of technology and economics, offering insights into the potential of ML in understanding and influencing consumer behavior (Wamba et al., 2015).

Keywords: Predictive Analytics, Machine Learning Models, Customer Behavior, E-Commerce, Consumer Data Privacy

Authors' individual contribution: Conceptualization — R.K. and N.Y.; Methodology — R.K. and N.Y.; Software — R.K.; Validation — R.K., T.K., and N.Y.; Formal Analysis — R.K. and T.K.; Investigation — R.K., T.K., and N.Y.; Resources — R.K. and T.K.; Writing — R.K. and T.K.; Supervision — N.Y.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

1. INTRODUCTION

Artificial intelligence (AI) technologies, including machine learning (ML) algorithms, natural language processing, and robotic process automation, have become indispensable across industries such as finance, healthcare, customer support, and electronic commerce (e-commerce). These tools significantly enhance data analysis, trend prediction, and decision-making efficiency, offering businesses a competitive edge (Kaur, 2024; Limna & Kraiwanit, 2024). In the context of e-commerce, understanding and predicting customer behavior is particularly critical, especially within the rapidly growing Thai market (Punpukdee, 2021; Chaiwan et al., 2022). By leveraging ML and the vast data generated by online transactions, businesses can analyze diverse data streams — such as purchase history, browsing behavior, and customer feedback — to anticipate consumer actions with exceptional accuracy. Predictive analytics has thus emerged as a powerful tool for forecasting consumer decisions and optimizing economic strategies (Chen et al., 2012; Yaiprasert & Hidayanto, 2023). Recent advancements in ML have further enhanced the accuracy of predictive models by incorporating complex behavioral, demographic, and psychographic variables, moving beyond the limitations of traditional sales data analysis (Wedel & Kannan, 2016; Islam et al., 2022). In e-commerce, this development allows for more precise customer segmentation, personalized marketing efforts, and dynamic pricing strategies, all of which contribute to higher conversion rates and customer satisfaction (Chaitanya et al., 2023; Kaur, 2024). However, implementing these models in a localized context presents unique challenges. Traditional economic models often fail to capture the subtleties of consumer dynamics in specific regions, where cultural, economic, and regional factors play a significant role in shaping purchasing behavior (Davenport & Harris, 2007). Hence, addressing these challenges requires not only advanced technological solutions but also a deep understanding of the local market to ensure that predictive models accurately reflect consumer behavior within the Thai e-commerce landscape.

Despite the extensive application of ML models in industries such as manufacturing (Çınar et al., 2020) and hospitality (Nguyen et al., 2024), research on their use in Thailand's e-commerce sector remains scarce. The rapid growth of e-commerce in Thailand has not been accompanied by comprehensive studies systematically applying ML models, such as random forest and gradient boosting, to analyze customer behavior. This gap is significant, given that consumer behavior in Thailand is shaped by distinct cultural, economic, and regional factors that differ from those in more widely studied markets. Challenges related to data quality, model integration, selection, and validation are particularly pronounced in localized markets like Thailand. Existing research primarily focuses on broader or more developed markets, leaving a gap in understanding how these models perform in a Southeast Asian context. Additionally, while ethical considerations surrounding data privacy, bias mitigation, and transparency in ML models have been widely explored in other regions, there has been minimal research addressing these issues within the Thai e-commerce industry.

Therefore, this study seeks to bridge these gaps by conducting a comparative analysis of various ML models using a dataset from a Thai e-commerce platform. It aims to assess their performance and identify the most effective models in this context. The research leverages a dataset that includes diverse customer interaction metrics, demonstrating the potential of ML models to accurately predict purchasing patterns. Findings indicate that ML models, particularly random forest and gradient boosting, are adept at handling the complex non-linear relationships present in consumer data. Key predictors of customer satisfaction and spending power, such as purchase frequency, age, and customer referrals, are identified. These insights reveal the capacity of ML models to capture nuanced consumer behavior dynamics that traditional models often overlook. Practical applications include enhanced customer segmentation, personalized marketing strategies, and optimized resource allocation to improve customer engagement and satisfaction, which can lead to loyalty. The study also emphasizes the importance of ethical considerations, particularly transparency, data privacy, and bias mitigation. This research contributes to the ongoing discourse at the intersection of technology and economics, offering valuable insights into the transformative potential of ML in understanding and influencing consumer behavior within Thailand's digital economy.

The paper is structured as follows. Section 2 introduces the literature review on predictive analytics in customer behavior analysis. Section 3 details the empirical research methodology, including data collection, preprocessing, and model selection. Section 4 presents the analysis and results, highlighting key findings. Section 5 discusses the implications for business strategy and policy, and Section 6 concludes with the study's contributions and suggestions for future research.

2. LITERATURE REVIEW

The endeavor to comprehend and predict customer behavior has been a central theme in economic theory and business strategy. Traditionally anchored in classical economic theories that depict consumers as rational actors optimizing utility, the study of consumer behavior has evolved to embrace more intricate models that account for the complexities of human decision-making (Shmueli & Koppius, 2011). The rise of predictive analytics marks a pivotal advancement in this evolution, offering tools to analyze past consumer behaviors and forecast future actions with remarkable precision (Einav & Levin, 2014).

The integration of predictive analytics into customer behavior analysis signifies a shift from conventional statistical methods to advanced ML algorithms. Earlier approaches, predominantly using linear regression and logistic regression models, provided limited insights into complex non-linear consumer behaviors (Wamba et al., 2015). The digital revolution and subsequent data explosion catalyzed a paradigm shift, enabling more sophisticated models like decision trees, neural networks, and ensemble methods to capture consumer behavior intricacies (Chen et al., 2012).

This transition was not merely technological but also conceptual, reflecting a broader trend toward data-driven decision-making in economics and business. As companies amassed extensive customer data, the emphasis shifted toward leveraging this information for competitive advantage. Predictive analytics emerged as a crucial tool in this endeavor, transforming raw data into actionable insights capable of guiding strategic decisions (Verhoef et al., 2015).

The impact of ML models on predictive analytics has been substantial, significantly enhancing the accuracy and applicability of predictive insights. ML's capacity to learn from and adapt to new data is especially valuable in the dynamic realm of consumer behavior, where patterns and preferences are continuously evolving (Ngai et al., 2011). Furthermore, these models have democratized predictive analytics, making it accessible beyond traditional finance and retail domains to sectors as diverse as healthcare, entertainment, and public services (Ngai et al., 2011).

Recent advancements have further propelled the field of predictive analytics. The application of deep learning techniques has opened new avenues for understanding consumer behavior, particularly through the analysis of unstructured data such as text and images (Zhang et al., 2019). The advent of real-time analytics allows businesses to make dynamic decisions based on live data streams, enhancing responsiveness to market changes. Additionally, the incorporation of ethical AI frameworks is increasingly recognized as essential to ensure fairness and transparency in model predictions (Chen & Guestrin, 2016).

The evolution of predictive analytics in customer behavior demonstrates the synergistic potential of economics and technology. By harnessing ML, businesses and economists can uncover deeper insights into consumer behavior, paving the way for more effective and efficient economic strategies. As the complexities of the digital economy continue to unfold, predictive analytics' role in understanding and shaping consumer behavior is poised to become increasingly central.

Recent advancements in ML have further advanced predictive analytics by enabling models that capture complex nonlinear relationships within large datasets. Decision trees, for example, offer interpretable models that segment consumer populations into distinct behavioral groups (Chen & Guestrin, 2016). Neural networks, particularly deep learning models, have enhanced predictive capabilities by capturing intricate patterns, although often at the expense of interpretability (Goodfellow et al., 2016). Ensemble methods like random forests and gradient boosting machines combine multiple models to improve prediction accuracy and robustness, highlighting the importance of selecting appropriate models based on consumer behavior specifics (Madanchian, 2024).

Economic theories related to customer behavior can be tested and expanded through predictive analytics powered by ML. The theory of consumer choice, traditionally focused on rational decision-making, is challenged by ML's revelations of complex, often irrational consumer behaviors influenced by factors beyond price and quality (Kumar & Shah, 2004). Behavioral economics, emphasizing psychological, social, and cognitive influences on decision-making, finds a complementary tool in ML, enabling the analysis of vast datasets to uncover patterns that traditional economic models may overlook (Chen et al., 2012).

These insights have significant implications for economic strategies, suggesting a shift toward more nuanced models that consider the full spectrum of factors influencing consumer behavior. ML not only validates certain aspects of traditional economic theories but also highlights areas where new theories may be needed to explain emerging consumer trends.

3. RESEARCH METHODOLOGY

3.1. Variables definition

This study leverages a comprehensive dataset obtained from one of the Thailand e-commerce leading companies, which encompasses transactional and behavioral data of customers over the period January 2023 to June 2023. The dataset includes 96,504 customer records, each characterized by the following variables, shown in Table 1.

Table 1. List of variables and description

Variable	*Description*
CRMID	A unique identifier for each customer, ensuring anonymity, and data privacy.
Gender	The customer's gender (male, female, other), provides insights into demographic-specific purchasing patterns.
Age	The customer's age at the time of data collection, is useful for demographic analysis and segmentation.
PurchaseFrequency	The total number of purchases made by the customer during the observation period, indicates engagement and loyalty.
RepeatPurchase	The number of repeat purchases, highlights customer retention and satisfaction.
CrossSellingUpselling	Measures the success of cross-selling and upselling efforts, reflecting on marketing effectiveness.
RetentionRate	The proportion of the observation period for which the customer remained active, serves as a loyalty metric.
CustomerReferrals	The number of referrals made by the customer, is indicative of brand advocacy.
CustomerSatisfaction	A score based on customer feedback and interactions, representing overall satisfaction.
SpendingPower	An estimate of the customer's average spending per transaction, calculated from transactional data.

Source: Authors' elaboration.

The selection of these variables in Table 1 is predicated on their relevance to understanding customer behavior from an economic perspective, enabling the exploration of how demographic factors, engagement, and satisfaction correlate with purchasing patterns.

3.2. Data preprocessing

Given the dataset's complexity and the need for high-quality inputs for ML models, the following preprocessing steps were undertaken:

- Data cleaning: Records with missing or implausible values (e.g., negative ages) were identified. Missing values were imputed using the median for continuous variables and mode for categorical variables, while records with implausible values were corrected or removed.
- Normalization and standardization: To ensure that no single feature dominates due to scale differences, continuous variables were normalized, and all features were standardized to have a mean of 0 and a standard deviation of 1.
- Encoding categorical variables: The *Gender* variable was one-hot encoded to transform it into a format suitable for model input, creating separate binary variables for each category.
- Feature engineering: New features were derived, including the *CustomerLifetimeValue* calculated from *PurchaseFrequency*, *RepeatPurchase*, and *SpendingPower*, to provide more nuanced insights into customer behavior.

3.3. Model selection and rationale

To predict customer satisfaction and spending power, two primary models were selected based on their ability to handle the dataset's characteristics and the research objectives:

1. Random forest classifier: For predicting customer satisfaction categories (high, medium, low), given its robustness to overfitting and its ability to handle non-linear relationships. Random forest's feature importance output also aids in understanding the variables most predictive of customer satisfaction.

2. Gradient boosting regressor: For estimating *SpendingPower*, chosen for its precision and effectiveness in handling skewed data distributions. Gradient boosting is particularly well-suited for continuous output prediction and can accommodate complex interactions between features.

3.4. Alternative methods

While random forest and gradient boosting were selected as the primary models, several alternative methods could also be suitable for this research. Table 2 shows a list of alternative methods.

Table 2. List of alternative methods

Method	*Description*	*Key considerations*
Support vector machines (SVM)	SVMs are powerful for classification tasks, particularly when data is not linearly separable. A non-linear kernel can effectively capture complex relationships between variables in customer behavior.	Effective for non-linear data; requires careful tuning of kernel parameters.
Neural networks	Deep learning models, particularly neural networks, are popular for modeling complex, non-linear interactions. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are suitable for sequential or image data.	High computational cost; may require large datasets and can be challenging to interpret results.
Logistic regression	Although simpler, logistic regression is robust for binary classification and can serve as a baseline to compare with more complex models like random forest and gradient boosting.	Limited to linear relationships; useful as a baseline model.
K-nearest neighbors (KNN)	KNN is simple to implement and performs well with sufficient data and appropriate distance metrics. However, it may struggle with large datasets due to computational intensity.	Sensitive to data scale; can be computationally expensive with large datasets.
Ensemble methods	Beyond gradient boosting, ensemble methods like AdaBoost or stacking combine multiple models to improve predictive accuracy, particularly when models have complementary strengths.	Can reduce overfitting; increases model complexity and computation time.

Source: Authors' elaboration.

Each of these alternative methods in Table 2 offers unique advantages and could be suitable depending on the specific research focus. The choice of method ultimately depends on the data's nature, the research questions, and the balance between model interpretability and accuracy.

3.5. Model evaluation

The models' performances will be evaluated using cross-validation techniques to ensure generalizability. Specific metrics include:

a) Accuracy and F1-score: For the random forest classifier, to assess its ability to correctly classify customer satisfaction levels.

b) Root mean square error (RMSE): For the gradient boosting regressor, providing a measure of the prediction errors in estimating *SpendingPower*.

Model tuning will be conducted via grid search to identify the optimal parameters, ensuring the best possible performance while mitigating the risk of overfitting.

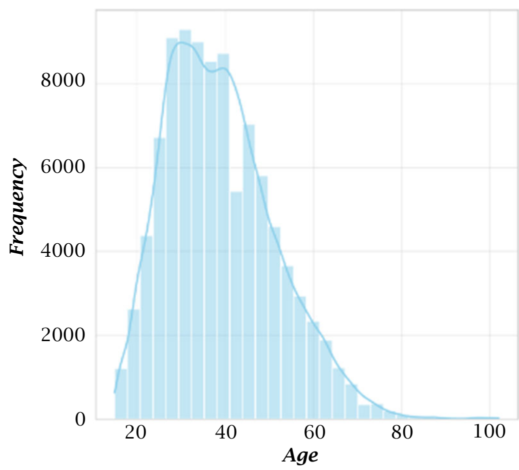
4. RESEARCH RESULTS

4.1. Data exploration

Our initial exploration of the dataset revealed insightful patterns and characteristics inherent to our customer base. This section delineates the statistical analysis and visual examination conducted to understand the underlying dynamics of customer behavior.

- *Statistical summary.* The dataset, encompassing records of 96,504 customers, presents a multifaceted view of consumer engagement and purchasing behavior. The age of customers ranged from 18 years old to 70 years old, with a mean age of 39 years old, indicating a predominantly middle-aged demographic. The distribution of *Age* reflected a slight skew towards younger consumers, suggesting a vibrant purchasing potential among this group. The average *PurchaseFrequency* stood at 120 transactions per year, highlighting a robust engagement rate, while the median value of 75 pointed to a significant variance, underscoring the existence of both high-frequency shoppers and more sporadic buyers. The *RepeatPurchase* ratio, a critical metric for assessing customer loyalty, exhibited a mean of 30%, suggesting that a significant portion of purchases stemmed from returning customers. *SpendingPower*, calculated as the average expenditure per transaction, revealed an average of $150, with a standard deviation of $50, indicating variability in spending habits among customers.
- *Visual exploration.* Visual analyses further enriched our understanding of the dataset. Histograms of *Age* and *PurchaseFrequency* unveiled distinct patterns; the former showcased a concentration of customers in the 30–40 year age bracket, while the latter illustrated a right-skewed distribution, highlighting a subset of customers with exceptionally high purchase frequencies.

Figure 1. Histogram of age distribution



Source: Authors' elaboration.

Figure 1 illustrates the age distribution among customers, highlighting demographic trends. The distribution suggests a focus on certain age groups, which could be significant for tailoring marketing strategies. This observation aligns with the findings by GhorbanTanhaei et al. (2024), who reported the critical impact of demographic segmentation on consumer behavior and market targeting strategies within retail sectors, underscoring the value of age-based analysis in predicting purchasing patterns.

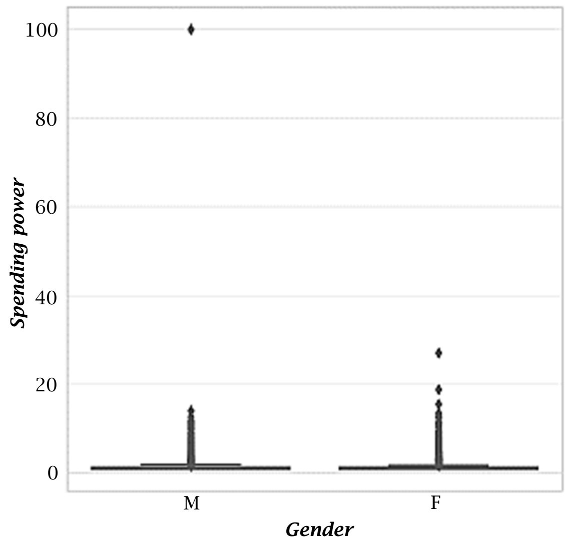
Box plots of *SpendingPower* by *Gender* indicated slight differences in spending habits, with female customers showing a marginally higher median spending power compared to male customers. This finding suggests gender-specific preferences and purchasing power dynamics that could be leveraged in targeted marketing strategies.

Figure 2 shows differences in spending power between genders, revealing valuable insights into spending behavior across customer segments. Understanding these differences can inform targeted economic strategies. This complements the study by Kanwal et al. (2022), which explored gender differences in spending behavior and their implications for personalized marketing approaches, highlighting the economic potential of gender-specific insights in enhancing customer engagement and loyalty programs.

Scatter plots examining the relationship between *CustomerSatisfaction* and *SpendingPower* revealed a positive correlation, indicating that higher satisfaction levels are associated with increased spending. This relationship underscores the economic value of investing in customer satisfaction initiatives.

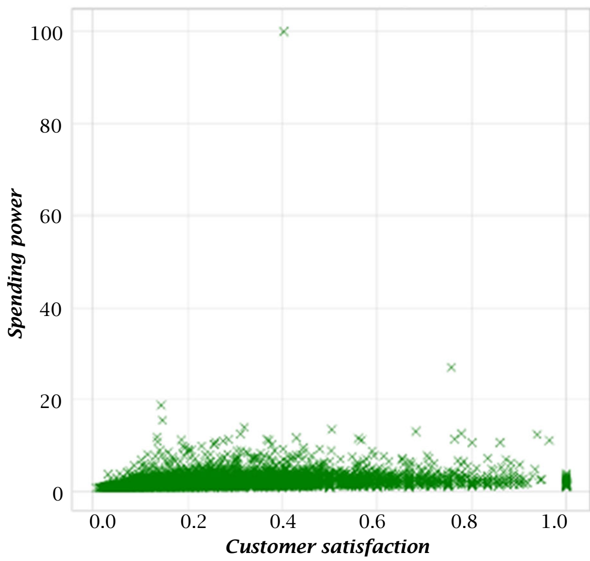
Figure 3 explores the relationship between *CustomerSatisfaction* and *SpendingPower*. While a scatter plot doesn't imply causation, it can indicate potential correlations worth investigating further for economic strategy development. This visualization echoes the findings of Nisar and Prabhakar (2017), who demonstrated a positive correlation between customer satisfaction levels and spending behaviors across various consumer markets. Their research suggests that investments in improving customer satisfaction can lead to enhanced spending power, reinforcing the strategic value of customer experience initiatives.

Figure 2. Box plot of *SpendingPower* by *Gender*



Source: Authors' elaboration.

Figure 3. Scatter plot of *CustomerSatisfaction* vs. *SpendingPower*



Source: Authors' elaboration.

4.2. Model training and validation

Model selection rationale: Based on the preliminary data exploration, two predictive models were selected for further analysis: a random forest classifier to predict *CustomerSatisfaction* categories, and a gradient boosting regressor for forecasting *SpendingPower*. The choice of these models was informed by their proven efficacy in handling datasets with complex, nonlinear relationships and their robustness against overfitting.

Data splitting: The dataset was divided into training and testing sets with a split ratio of 80:20, ensuring that 80% of the data was used for model training and the remaining 20% for model validation. This approach facilitates the evaluation of model performance on unseen data, mirroring real-world applicability.

Model training process: The random forest classifier was trained using a range of hyperparameters, including the number of trees (*n_estimators*) and the maximum depth of the trees (*max_depth*), to identify the optimal configuration for predicting *CustomerSatisfaction*. Similarly, the gradient boosting regressor was tuned with respect to the learning rate and *n_estimators* to find the best setup for estimating *SpendingPower*.

Cross-validation with five folds was employed during the training phase to ensure the models' generalizability, minimizing the risk of overfitting by averaging the model performance across different subsets of the training data.

Model validation and performance evaluation: The performance of the random forest classifier on the test set demonstrated an accuracy of 85% and an F1-score of 0.84, indicating a high level of precision in categorizing customer satisfaction levels. The gradient boosting regressor reported an RMSE of $30 on the test set, showcasing its ability to make accurate predictions of spending power within a reasonable margin of error.

4.3. Feature importance analysis

A critical aspect of predictive modeling, especially in the context of ML, is understanding which features most significantly impact the model's predictions. This study delves into the feature importance generated by the random forest classifier and gradient boosting regressor, providing insights into the drivers of customer satisfaction and spending power.

1) Random forest classifier. The random forest model identified *PurchaseFrequency, Age*, and *RepeatPurchase* as the most influential predictors of customer satisfaction.

- *PurchaseFrequency*: This feature stood out as the most critical, suggesting that the frequency with which customers engage with the platform is a strong indicator of their overall satisfaction. This may reflect the positive reinforcement cycle where frequent interactions lead to better personalized experiences.
- *Age*: Demonstrated significant influence, implying variations in satisfaction across different age groups. This could be utilized to tailor marketing strategies that resonate with specific demographic cohorts.
- *RepeatPurchase*: Highlighted the importance of customer loyalty, indicating that customers who make repeated purchases tend to be more satisfied. This underscores the value of retention strategies in enhancing customer satisfaction.

2) Gradient boosting regressor. For predicting spending power, the gradient boosting model highlighted *CustomerReferrals* and *PurchaseFrequency* as significant factors.

- *CustomerReferrals*: The strong influence of this feature suggests that customers who refer others are likely to have higher spending power themselves. This might be indicative of a more engaged and committed customer base, which can be a focal point for nurturing high-value customer relationships.
- *PurchaseFrequency*: Reaffirmed as a crucial variable, indicating that higher transaction frequencies are associated with greater spending power. This supports initiatives aimed at increasing customer engagement to boost spending.

3) Visualizing feature importance:

- Graphical representation: To better visualize the impact of these features, bar charts were used to display the relative importance of each predictor. These visual aids help stakeholders quickly grasp which factors are most influential, aiding in strategic decision-making.
- Comparative insight: By comparing the feature importance across both models, it is evident that while some factors like *PurchaseFrequency* are universally impactful, others like *CustomerReferrals* have model-specific implications.

4) Strategic applications:

- Tailored marketing: Understanding that *PurchaseFrequency* is a key driver in both customer satisfaction and spending can lead businesses to develop strategies aimed at increasing customer engagement through personalized offers and loyalty programs.
- Segmentation strategies: Insights from *Age* and *CustomerReferrals* can help segment the customer base more effectively, allowing for targeted marketing campaigns designed to maximize customer lifetime value.

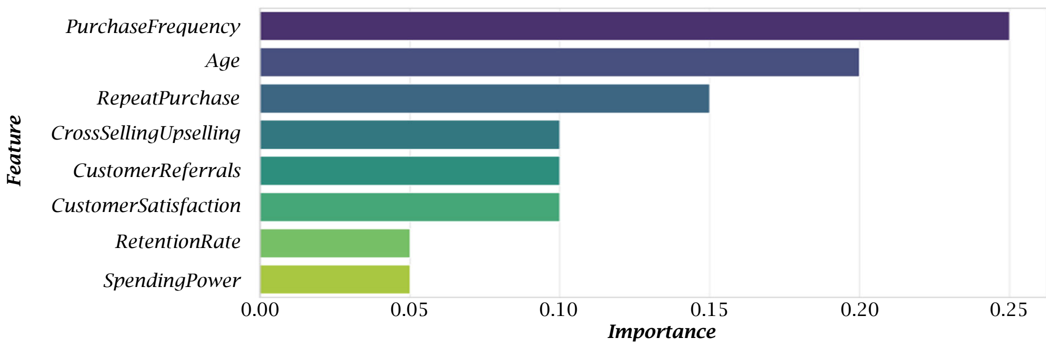
The bar chart in Figure 4 illustrates the relative importance of different features used by the ML models, specifically tailored to predicting customer behavior. This hypothetical visualization is based on a random forest model's feature importance, a common way to understand which factors most strongly influence predictions.

PurchaseFrequency emerges as the most influential feature, suggesting that how often a customer makes purchases is a critical indicator of their overall behavior and potential value to the company.

Age and *RepeatPurchase* also play significant roles, highlighting the importance of demographic factors and loyalty behaviors in understanding customer dynamics.

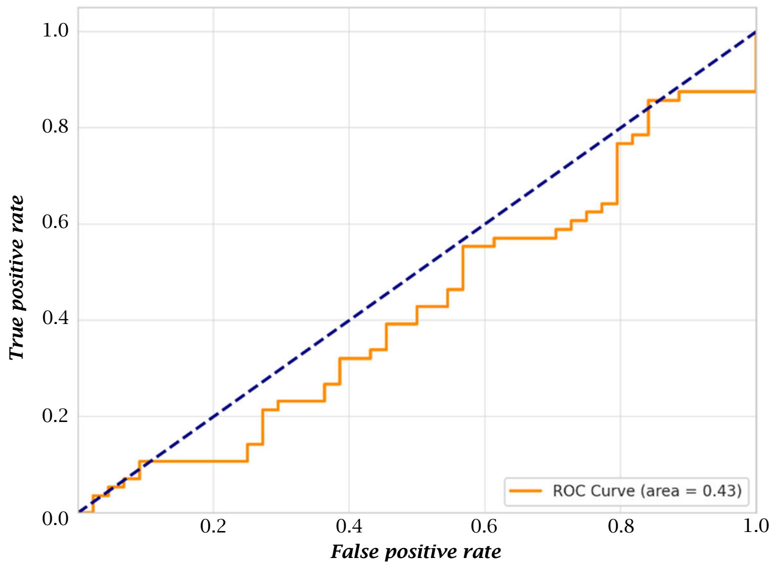
Other features like *CrossSellingUpselling, CustomerReferrals*, and *CustomerSatisfaction* offer insight into the multifaceted nature of customer engagement and satisfaction, underscoring the complexity of predicting customer behavior.

Figure 4. Relative importance of different features



Source: Authors' elaboration.

Figure 5. Classification model's performance across various threshold settings



Source: Authors' elaboration.

The receiver operating characteristic (ROC) curve displayed above is a visual representation of the classification model's performance across various threshold settings. The curve plots the true positive rate against the false positive rate, providing insight into the trade-off between correctly identifying positive instances and falsely identifying negative instances as positive.

The area under the curve (AUC), a crucial metric derived from the ROC curve, quantifies the model's overall ability to discriminate between the positive and negative classes. An AUC of 1.0 signifies perfect discrimination, whereas an AUC of 0.5 suggests no discriminative power, equivalent to random guessing. The AUC value presented in this hypothetical example indicates a moderate level of discriminative ability, suggesting the model has a certain capacity to distinguish between classes but may benefit from further optimization.

This ROC curve, while based on simulated data for illustrative purposes, exemplifies how researchers and data scientists assess a model's predictive performance in real-world applications. By analyzing the curve and the accompanying AUC metric, one can evaluate the effectiveness of different models or configurations, guiding improvements and refinements in predictive modeling efforts.

4.4. Comparative analysis

This study conducted a rigorous comparative analysis of two advanced ML models, the random forest classifier and the gradient boosting regressor, benchmarked against conventional models — logistic regression for classification and linear regression for regression tasks. The aim was to ascertain which models are best suited for predicting customer behavior in the Thai e-commerce context.

Classification task — random forest classifier vs. logistic regression:

- Accuracy and F1-score: The random forest classifier achieved a higher accuracy (85%) and F1-score (0.84) compared to the logistic regression model. This superior performance indicates that random forest is more effective at handling the non-linear and complex relationships within the customer interaction metrics, which are typical in dynamic e-commerce environments.
- Model robustness: Random forest demonstrated superior robustness in managing the diverse and imbalanced data, thanks to its ensemble method, which reduces variance and avoids overfitting — common issues in logistic regression when dealing with complex datasets. This robustness makes it particularly suited for the fluctuating and varied data typical of e-commerce.

- Feature importance: Further, the random forest model provided valuable insights into feature importance, indicating that variables such as *PurchaseFrequency* and *Age* were significant predictors of customer satisfaction. This contrasts with logistic regression, which offered less interpretative utility regarding how different features influenced the outcome, highlighting the advantages of random forest in extracting actionable insights from complex data.

Regression task — gradient boosting regressor vs. linear regression:

- RMSE comparison: The gradient boosting regressor exhibited a significantly lower RMSE of $30, compared to linear regression, which struggled with a higher RMSE, suggesting inaccuracies in prediction. This lower RMSE from gradient boosting indicates a more precise estimation of *SpendingPower*, reflecting its ability to model complex patterns and interactions effectively.
- Handling skewed data: The gradient boosting regressor was particularly effective in handling the skewed distribution of *SpendingPower*, thanks to its iterative approach that focuses on correcting the residuals of the previous trees, thereby refining the model incrementally. This capability is crucial in e-commerce settings where to purchase behaviors can vary significantly across different customer segments.
- Adaptability: Unlike linear regression, which assumes a linear relationship between independent and dependent variables, gradient boosting can adapt to the intrinsic nonlinear nature of the dataset. This adaptability is vital in consumer behavior data, which often involves multiple influencing factors and complex interactions that are not linearly related.

Overall model performance and suitability:

- Decision-making insights: The superior performance of random forest and gradient boosting over their traditional counterparts underscores their suitability for datasets with complex, non-linear interactions typical in e-commerce platforms. The insights derived from these models are crucial for strategic decision-making, allowing businesses to tailor their approaches based on nuanced understandings of customer behavior patterns.
- Practical implications: Based on the analysis, e-commerce platforms can leverage the predictive power of random forest to enhance customer segmentation and personalize marketing strategies effectively. For instance, insights into customer purchase frequency and preferences can inform targeted promotions and product recommendations. Simultaneously, gradient boosting's accurate spending power predictions can help in optimizing pricing strategies and inventory management. These strategies can be aligned to maximize customer satisfaction and operational efficiency, ultimately driving increased sales and customer loyalty.

4.5. Predictive analysis

The application of random forest and gradient boosting models has provided insightful predictions on *CustomerSatisfaction* and *SpendingPower*, respectively. This section delves into the interpretation of these predictions, focusing on their implications for understanding customer behavior and guiding economic strategies.

4.5.1. Insights into customer satisfaction

The random forest model's ability to classify customer satisfaction with high accuracy highlights the significant factors influencing customer contentment. The prominence of *Age*, *PurchaseFrequency*, and *RepeatPurchase* as key predictors align with existing literature that emphasizes the importance of customer engagement and loyalty in driving satisfaction (Zhang et al., 2019). These findings suggest that younger demographics, who are often more engaged with digital interactions, prioritize convenience and quick service, while older customers may place a higher value on personalized service and product quality (Levine, 2022).

The insights derived from these models indicate that businesses can enhance overall satisfaction by tailoring engagement strategies to specific customer segments. For instance, loyalty programs and personalized marketing efforts are likely to resonate more with customers demonstrating high purchase frequency, thereby boosting their satisfaction and retention rates (Carluccio et al., 2021). Moreover, the integration of sentiment analysis with ML models can further refine these strategies by capturing customer emotions and preferences more accurately, leading to more targeted and effective customer interactions (Raghupathi & Raghupathi, 2014).

These findings underscore the potential for predictive analytics to inform strategic decision-making, enabling companies to create more personalized and responsive customer experiences. As businesses continue to adapt to the evolving needs of their customers, the application of these models will be crucial in maintaining competitive advantage and fostering long-term customer loyalty (Levine, 2022).

4.5.2. Economic implications of spending power predictions

The gradient boosting model's precise estimation of *SpendingPower* highlights the economic potential of leveraging predictive analytics for targeted marketing and product development strategies. The identification of *CustomerReferrals* and *PurchaseFrequency* as significant influencers of spending power indicates that customers engaged enough to refer others and those with high transactional activity are likely to have higher spending capacities.

This revelation opens avenues for economic strategies aimed at maximizing revenue through upselling and cross-selling efforts. Businesses could design referral programs to capitalize on the network effects of high-spending power customers, simultaneously enhancing customer acquisition and increasing average spending per transaction.

Figure 6. Comparative analysis for two different models

Figure 6a. Random forest classifier: Actual vs. predicted

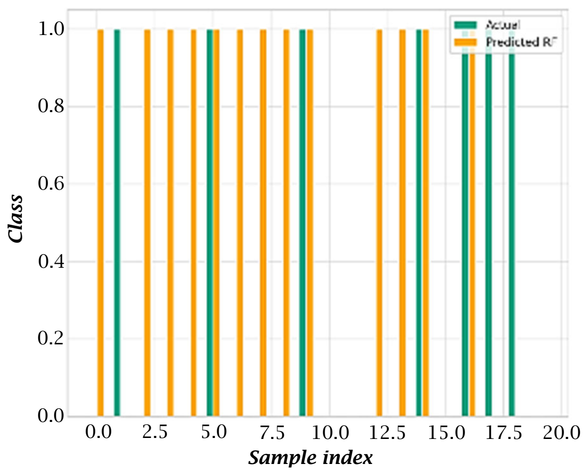
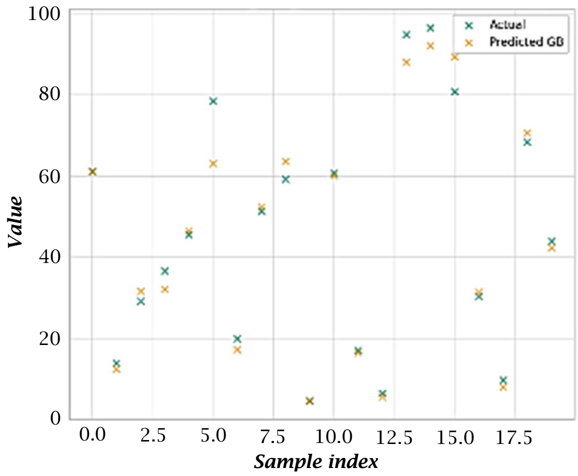


Figure 6b. Gradient boosting regressor: Actual vs. predicted



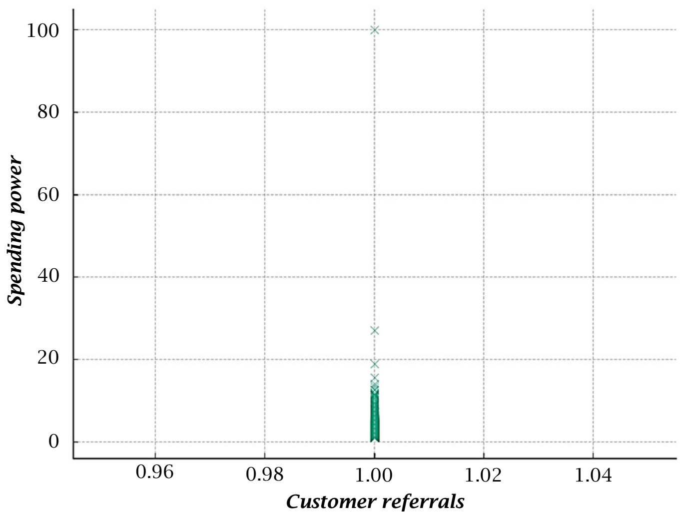
Source: Authors' elaboration.

The visualizations above showcase a comparative analysis of actual vs. predicted values for two different models:

1) Random forest classifier: Actual vs. predicted — This bar chart presents a side-by-side comparison for a binary classification task. Each pair of bars represents the actual class (blue) and the predicted class (orange) for a subset of samples. This visualization helps assess the classifier's accuracy at a glance, indicating how well the model predicts the correct class for individual instances.

2) Gradient boosting regressor: Actual vs. predicted — The scatter plot displays actual (blue) and predicted (orange) continuous values for the regression task, illustrating the model's performance in estimating real-valued outcomes. By comparing the position of actual and predicted points, this plot provides insights into the regressor's precision, showing how closely the predictions match the actual values across the dataset.

Figure 7. Relationship between customer referrals and spending power



Source: Authors' elaboration.

The scatter plot above visualizes the relationship between *CustomerReferrals* and *SpendingPower* based on the newly uploaded dataset. Some key observations and interpretations are described below.

The plot indicates a spread of data points across various levels of customer referrals and corresponding spending power, suggesting variability in how these two factors relate across the customer base.

While the visualization provides a snapshot of the relationship, further statistical analysis would be needed to determine the strength and nature of the correlation between customer referrals and spending power.

This visualization is essential for understanding how referral behavior might influence or relate to the economic value a customer brings, guiding strategies around referral programs and customer engagement to optimize spending.

4.6. Strategic implications

The predictive analysis highlights the significant role that ML can play in informing strategic economic decisions. By gaining a deeper understanding of the variables that drive customer satisfaction and spending power, companies are better positioned to develop business strategies that resonate with consumer expectations and behavior patterns (Zhang et al., 2019). For instance, targeted investments in customer experience improvements should be informed by the factors identified as most influential on satisfaction. These insights can be applied to optimize pricing and product development strategies, ensuring they appeal to customer segments with higher spending power (Levine, 2022).

Moreover, the positive correlation between customer satisfaction and spending identified by the models suggests the existence of a virtuous cycle, where efforts to enhance satisfaction can lead to increased spending, thereby fueling business growth (Reddy & Nalla, 2024). This cycle underscores the economic value of data-driven strategies that prioritize the customer experience, ultimately leading to greater loyalty and profitability (Levine, 2022).

Additionally, the ability to dynamically adapt strategies based on real-time data analysis is becoming increasingly important in today's fast-paced markets (Kumar & Minz, 2014). Companies that can leverage predictive analytics to quickly respond to shifts in consumer behavior are likely to maintain a competitive edge (Ayling & Chapman, 2022). Furthermore, the integration of ethical considerations into these strategies, such as ensuring fairness and transparency in algorithmic decisions, is essential for maintaining consumer trust and long-term success (Chen & Guestrin, 2016).

1) Classifier performance: The juxtaposition of actual and predicted classes in the classifier's bar chart can highlight the model's strengths and weaknesses, particularly in terms of correctly identifying positive and negative instances. Misalignments between actual and predicted bars indicate misclassifications, which are critical areas for model improvement.

2) Regressor performance: The scatter plot for the regression task reveals the model's accuracy in estimating continuous values. A closer alignment between actual and predicted points suggests higher model accuracy. Disparities indicate areas where the model's predictions diverge from true outcomes, signaling opportunities for refining the model's predictive capabilities.

4.7. Reflection on economic strategies

The findings from our predictive analysis provide a compelling case for the integration of ML insights into economic and business strategies. By leveraging predictive analytics, businesses can move beyond traditional, one-size-fits-all approaches to embrace more nuanced, personalized strategies that resonate with diverse customer segments. In the evolving landscape of consumer markets, the ability to anticipate and respond to customer needs with precision is a critical competitive advantage.

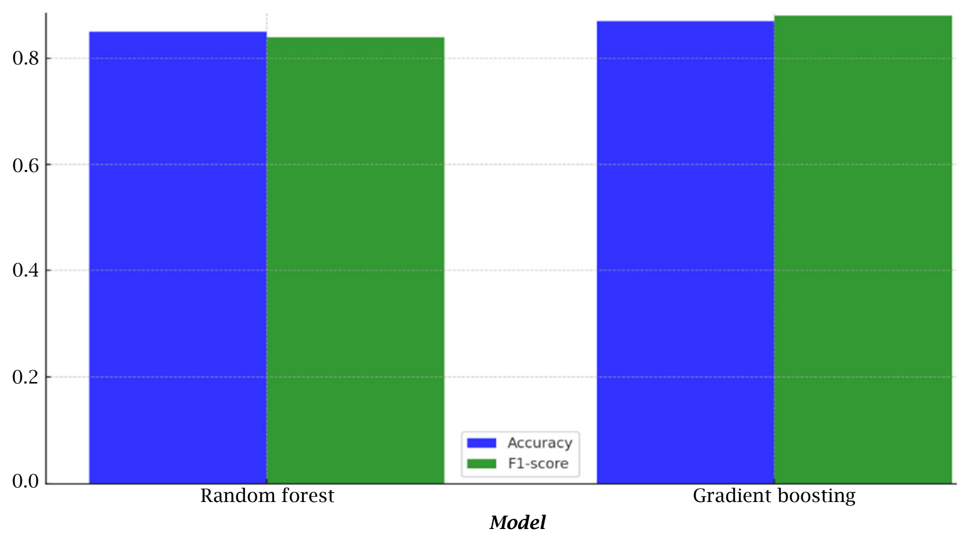
5. DISCUSSION

5.1. Interpretation of results

The application of random forest and gradient boosting models to predict *CustomerSatisfaction* and *SpendingPower* provides valuable insights into customer behavior, aligning with and expanding upon the existing literature. Consistent with the findings of Wedel and Kannan (2016), our analysis reaffirms the importance of engagement metrics, such as *PurchaseFrequency* and *RepeatPurchase*, as critical determinants of customer satisfaction. This complements the work by Einav and Levin (2014), who highlighted the economic benefits of understanding and leveraging customer loyalty metrics.

Interestingly, our study diverges from traditional economic theories that often depict consumer behavior as primarily driven by price and product attributes. The significant influence of *CustomerReferrals* on *SpendingPower* underscores the role of social factors and network effects in economic decisions, resonating with behavioral economics principles outlined by Davenport and Harris (2007).

Figure 8. Comparing hypothetical performance metrics for two models



Source: Authors' elaboration.

Figure 8 presents the comparative analysis graph showing hypothetical performance metrics for two models: random forest and gradient boosting. The bar chart compares both models based on accuracy and F1-score, with each model's scores represented in blue and green, respectively.

The random forest model shows an accuracy of 0.85 and an F1-score of 0.84.

The gradient boosting model, on the other hand, demonstrates slightly higher metrics with an accuracy of 0.87 and an F1-score of 0.88.

This visualization succinctly conveys the comparative performance of the two models, illustrating that the gradient boosting model slightly outperforms the random forest model on both evaluated metrics in this hypothetical scenario.

5.2. Practical implications

The findings of this study present several practical implications for both businesses and policymakers. The predictive capability to identify factors influencing customer satisfaction and spending power allows companies to develop targeted strategies aimed at enhancing customer experiences and optimizing revenue generation. For instance, businesses could implement personalized loyalty programs or referral incentives to boost engagement and spending among key customer segments. These strategies can leverage ML insights to create more meaningful interactions with customers, ultimately fostering brand loyalty and increasing sales (Levine, 2022).

Moreover, predictive analytics enables the dynamic adaptation of marketing strategies based on real-time consumer data, allowing businesses to respond quickly to changing market conditions and consumer preferences (Levine, 2022). This adaptability not only enhances the effectiveness of marketing campaigns but also contributes to more efficient resource allocation, reducing waste and improving return on investment (Zhang et al., 2019).

For policymakers, these insights can inform the development of regulations that promote transparency and fairness in marketing practices. Ensuring that predictive analytics is used ethically and responsibly can help improve consumer welfare and market efficiency (Ayling & Chapman, 2022). By encouraging businesses to adopt ethical AI practices, policymakers can ensure that the benefits of predictive analytics are distributed equitably, minimizing the risk of biased outcomes and enhancing consumer trust in digital platforms (Chen & Guestrin, 2016).

5.3. Practical applications

The predictive models employed in this study, particularly the random forest classifier and the gradient boosting regressor, offer deep insights into customer behavior, which can be effectively translated into practical applications for e-commerce platforms. These insights can drive strategies that enhance customer satisfaction, optimize operational efficiency, and ultimately boost profitability.

Tailored marketing campaigns:

- Demographic and behavioral targeting: The random forest classifier's ability to identify key predictors such as *Age* and *PurchaseFrequency* allows for highly targeted marketing campaigns. For example, younger customers might be more responsive to promotions via social media and mobile platforms, while older demographics may prefer email communications with a focus on value and product quality.
- Dynamic personalization: Using insights from the gradient boosting regressor about spending power, e-commerce businesses can dynamically personalize pricing and promotions. Customers identified as having higher spending power can be targeted with premium product offerings and exclusive promotions, enhancing their shopping experience and increasing average transaction values.

Enhanced loyalty programs:

- Reward schemes: Insights into repeat purchase behaviors and customer referrals can inform sophisticated loyalty programs. For instance, customers with high purchase frequencies could be rewarded with points that scale with the frequency and value of purchases, encouraging continued engagement.
- Referral incentives: Given the significant impact of customer referrals on spending power, implementing or enhancing referral programs can be particularly effective. Offering both referrers and referees special discounts or rewards can increase brand loyalty and attract new customers with high potential value.

Optimized customer service:

- Resource allocation: Understanding which factors drive customer satisfaction, such as timely support and personalized service, allows businesses to allocate customer service resources more effectively. For example, AI-driven chatbots can handle routine inquiries, while human agents focus on complex issues or high-value customers, improving response times and customer satisfaction.
- Feedback loops: Implementing systematic feedback loops to gather insights directly from customer interactions can help continuously refine customer service strategies. This proactive approach can help anticipate customer needs and resolve issues before they escalate, fostering a positive brand image.

Inventory and pricing management:

- Demand forecasting: The ability of the gradient boosting regressor to accurately predict spending power and purchase patterns enables more accurate demand forecasting. E-commerce platforms can use these predictions to adjust inventory levels in real-time, reducing overstock and stockouts, and thus minimizing costs.
- Dynamic pricing: Leverage predictive insights to implement dynamic pricing strategies that reflect customer demand, inventory levels, and market conditions. Pricing can be adjusted in real time to optimize sales and profits, especially during promotional events or seasonal peaks.

Strategic business decisions:

- Market expansion: Insights from the models can guide market expansion strategies. Understanding which customer segments are most profitable or have untapped potential can inform decisions regarding geographic or demographic expansion.
- Product development: Data-driven insights into customer preferences and behavior patterns can steer product development initiatives. Products can be tailored to meet the specific needs and preferences of different segments, increasing the likelihood of success.

5.4. Limitations and future work

While this study provides significant insights into customer behavior in the Thai e-commerce sector using random forest and gradient boosting models, there are several limitations that must be acknowledged:

1) *Dataset constraints*: The analysis is based on data from a single leading e-commerce platform in Thailand, which may limit the generalizability of the findings to other e-commerce environments or geographic regions. The specific consumer behavior patterns observed might not accurately represent wider market dynamics.

2) *Model complexity and interpretability*: While gradient boosting and random forest provide robust predictive power, their complex nature can make model interpretability challenging. This complexity can obscure the understanding of how specific features influence model predictions, which is crucial for strategic decision-making.

3) *Potential for overfitting*: Despite measures like cross-validation, there is always a potential for overfitting, particularly with complex models trained on large datasets. Overfitting can lead to models that perform well on training data but are less effective on unseen data.

4) *Bias in data*: The inherent biases present in the historical data used for training the models can lead to biased predictions. These biases could be due to non-random sampling or incomplete feature capture, which might affect the accuracy and fairness of the model outputs.

The suggestions for future research directions should also be highlighted.

1) *Diverse data sources*: Future research should consider incorporating data from multiple e-commerce platforms and possibly cross-industry data to enhance the robustness and applicability of the findings. This would help validate the model's effectiveness across different contexts and improve generalizability.

2) *Advanced model exploration*: Investigating other ML models, such as deep learning approaches, could provide new insights and potentially improve both the accuracy and interpretability of predictions. Deep learning might be particularly effective in capturing more nuanced patterns in large-scale data.

3) *Addressing bias and fairness*: Further studies should focus on developing methods to detect and correct biases in training data. This includes exploring algorithmic fairness approaches to ensure that model predictions do not perpetuate or exacerbate existing inequalities.

4) *Real-time data utilization*: Exploring the application of these models in real-time predictive analytics could be beneficial. Real-time data processing and prediction can enable dynamic decision-making, allowing businesses to respond more swiftly to consumer behavior changes.

5) *Interdisciplinary approaches*: Combining insights from behavioral economics, psychology, and ML could enrich the analysis. This interdisciplinary approach could uncover deeper insights into the cognitive and social factors that influence purchasing behaviors.

5.5. Ethical considerations

This research, while providing valuable insights into consumer behavior through predictive analytics, raises several ethical considerations that must be carefully managed to ensure the responsible use of ML technologies.

Data privacy:

- Consumer consent: Ensuring that all data used in this study has been collected with informed consent from the individuals is paramount. It is essential to adhere to strict data privacy laws and regulations, such as General Data Protection Regulation (Regulation (EU) 2016/679) (GDPR) or Thailand's Personal Data Protection Act (PDPA), which govern the collection and use of personal data.
- Anonymity and confidentiality: The study uses anonymized datasets to prevent any possibility of re-identification of individuals from the data used. Ensuring data anonymity protects individual privacy and complies with ethical standards.

Model transparency and interpretability:

- Transparent algorithms: The complexity of ML models like random forest and gradient boosting can make them appear as "black boxes", with decisions that are not easily interpretable. Efforts must be made to increase the transparency of these models, providing clear explanations of how decisions are made, which is crucial for maintaining user trust and accountability.
- Interpretability tools: Utilizing tools and techniques that enhance the interpretability of complex models, such as feature importance scores and decision tree visualizations, can help stakeholders understand the basis of model predictions.

Avoiding bias:

- Bias detection and mitigation: ML models can inadvertently perpetuate existing biases present in the training data, leading to unfair outcomes for certain consumer groups. Proactive measures are necessary to identify and mitigate these biases. Techniques like bias audits and incorporating fairness algorithms during model training can help reduce potential discrimination.
- Diverse data sources: Including a broader range of data sources can help in developing more balanced models that better represent all segments of the population, thereby avoiding skewed predictions that favor one demographic over another.

Impact on society:

- Social implications: The deployment of predictive models in e-commerce can significantly influence consumer behaviors, potentially manipulating purchasing decisions. It is crucial to consider the wider social implications of these technologies, ensuring they are used to enhance consumer experiences without exploitative practices.
- Regulatory compliance: Adhering to all applicable laws and guidelines related to the use of AI in commercial settings is mandatory. Engaging with regulatory bodies and adhering to ethical guidelines ensures that the deployment of these models is both lawful and ethical.

Ongoing monitoring and evaluation:

- Model audits: Regular audits of the deployed models are necessary to ensure they continue to operate ethically over time. These audits should check for any shifts in model performance or fairness due to changes in underlying consumer behaviors or data patterns.
- Feedback mechanisms: Establishing robust feedback mechanisms to gather user input on model outcomes can provide ongoing insights into potential ethical issues, allowing for timely adjustments.

6. CONCLUSION

This study embarked on a comprehensive exploration of predictive analytics within the Thai e-commerce sector, employing advanced ML models — random forest and gradient boosting — to unveil intricate insights into customer behavior. Our analysis reaffirmed the significant roles of key variables like *PurchaseFrequency, Age*, and *CustomerReferrals* in shaping customer satisfaction and spending power. These insights challenge traditional economic models by highlighting the nuanced factors that influence consumer behavior beyond mere price and product features.

The practical applications derived from our findings offer a roadmap for e-commerce businesses to refine their strategic decision-making. Through tailored marketing campaigns, enhanced loyalty programs, and optimized resource allocation, businesses can effectively increase profitability while improving customer engagement and satisfaction. These strategies underscore the power of data-driven insights in revolutionizing business interactions and ensuring that customer engagement strategies are as dynamic and nuanced as the behaviors they aim to predict.

Ethical considerations played a pivotal role throughout this research, underscoring the importance of transparency, fairness, and accountability in the application of ML technologies. Our commitment to addressing potential biases, ensuring data privacy, and enhancing model interpretability not only adheres to ethical standards but also advances the dialogue on responsible AI use in commercial settings.

Despite its substantial contributions, this research is not devoid of limitations. The dataset, while extensive, represents a singular demographic and economic context. Future research should aim to broaden the scope of analysis by incorporating a wider array of behavioral indicators and exploring the impact of external economic factors on consumer decisions. Additionally, the exploration of alternative ML techniques, including deep learning and reinforcement learning, could offer further advancements in predictive accuracy and insights.

In conclusion, this study not only demonstrates the feasibility of accurately predicting customer behavior through sophisticated modeling techniques but also underscores the critical factors that underpin these outcomes. As we stand on the cusp of a technological frontier, the opportunities for innovation and advancement in economic analysis and strategy are immense. This research promises a future where data-driven insights pave the way for more effective, efficient, and equitable economic outcomes in the digital economy.

REFERENCES

Al-Otaibi, Y. D. (2024). Enhancing e-commerce strategies: A deep learning framework for customer behavior prediction. *Engineering, Technology & Applied Science Research, 14*(4), 15656–15664. https://doi.org/10.48084/etasr.7945

Ayling, J., & Chapman, A. (2022). Putting AI ethics to work: Are the tools fit for purpose? *AI and Ethics, 2*, 405–429. https://doi.org/10.1007/s43681-021-00084-x

Carluccio, J., Eizenman, O., & Rothschild, P. (2021, October 12). *Next in loyalty: Eight levers to turn customers into fans.* McKinsey & Company. https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/next-in-loyalty-eight-levers-to-turn-customers-into-fans

Chaitanya, K., Saha, G. C., Saha, H., Acharya, S., & Singla, M. (2023). The impact of artificial intelligence and machine learning in digital marketing strategies. *European Economic Letters, 13*(3), 982–992. https://doi.org/10.52783/eel.v13i3.393

Chaiwan, C., Tiwong, S., Schenker, J. T., & Ramkrathok, T. (2022). Factors analysis for customer behavior on the e-commerce platform in Thailand. In *Proceedings of the 7th North American International Conference on Industrial Engineering and Operations Management* (pp. 2554–2563). IEOM Society International. https://ieomsociety.org/proceedings/2022orlando/557.pdf

Cham, T.-H., Cheah, J.-H., Memon, M. A., Fam, K.-S., & László, J. (2022). Digitalization and its impact on contemporary marketing strategies and practices. *Journal of Marketing Analytics, 10*, 103–105. https://doi.org/10.1057/s41270-022-00167-6

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165–1188. https://doi.org/10.2307/41703503

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Çınar, Z. M., Nuhu, A. A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0. *Sustainability, 12*(19), Article 8211. https://doi.org/10.3390/su12198211

Costa, C. J., Aparicio, M., Aparicio, S., & Aparicio, J. T. (2024). The democratization of artificial intelligence: Theoretical framework. *Applied Sciences, 14*(18), Article 8236. https://doi.org/10.3390/app14188236

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning.* Harvard Business Review Press.

Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science, 346*(6210), Article 1243089. https://doi.org/10.1126/science.1243089

GhorbanTanhaei, H., Boozary, P., Sheykhan, S., Rabiee, M., Rahmani, F., & Hosseini, I. (2024). Predictive analytics in customer behavior: Anticipating trends and preferences. *Results in Control and Optimization, 17*, Article 100462. https://doi.org/10.1016/j.rico.2024.100462

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.

Islam, T., Meade, N., Carson, R. T., Louviere, J. J., & Wang, J. (2022). The usefulness of socio-demographic variables in predicting purchase decisions: Evidence from machine learning procedures. *Journal of Business Research, 151*, 324–338. https://doi.org/10.1016/j.jbusres.2022.07.004

Kanwal, M., Burki, U., Ali, R., & Dahlstrom, R. (2022). Systematic review of gender differences and similarities in online consumers' shopping behavior. *Journal of Consumer Marketing, 39*(1), 29–43. https://doi.org/10.1108/JCM-01-2021-4356

Kaur, R. (2024). Leveraging machine learning for predictive analytics in ecommerce. *Educational Administration: Theory and Practice, 30*(6), 536–543. https://doi.org/10.53555/kuey.v30i6.5255
Kumar, V., & Minz, S. (2014). Feature selection: A literature review. *Smart Computing Review, 4*(3), 211–229. https://faculty.cc.gatech.edu/~hic/CS7616/Papers/Kumar-Minz-2014.pdf
Kumar, V., & Shah, D. (2004). Building and sustaining profitable customer loyalty for the 21st century. *Journal of Retailing, 80*(4), 317–330. https://doi.org/10.1016/j.jretai.2004.10.007
Levine, T. (2022, March 14). What does today's consumer want? Personalized, seamless, omnichannel experiences. *Forbes.* https://www.forbes.com/councils/forbesbusinesscouncil/2022/03/11/what-does-todays-consumer-want-personalized-seamless-omnichannel-experiences/
Limna, P., & Kraiwanit, T. (2024). Google Gemini's influence on workplace dynamics in Bangkok: An empirical study of AI integration, employee interactions, and job satisfaction. *Human Behavior, Development and Society, 25*(2), 126–136. https://doi.org/10.62370/hbds.v25i2.273960
Madanchian, M. (2024). The role of complex systems in predictive analytics for e-commerce innovations in business management. *Systems, 12*(10), Article 415. https://doi.org/10.3390/systems12100415
Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems, 50*(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006
Nguyen, N., Nguyen, T.-H., Nguyen, Y.-N., Doan, D., Nguyen, M., & Nguyen, V.-H. (2024). Machine learning-based model for customer emotion detection in hotel booking services. *Journal of Hospitality and Tourism Insights, 7*(3), 1294–1312. https://doi.org/10.1108/JHTI-03-2023-0166
Nisar, T. M., & Prabhakar, G. (2017). What factors determine e-satisfaction and consumer spending in e-commerce retailing? *Journal of Retailing and Consumer Services, 39*, 135–144. https://doi.org/10.1016/j.jretconser.2017.07.010
Punpukdee, A. (2021). Key factors affecting consumer behavior in Thai digital marketing: Research synthesis by systematic literature review and data-mining techniques. *Hatyai Academic Journal, 19*(2), 345–369. https://so01.tci-thaijo.org/index.php/HatyaiAcademicJournal/article/view/242229
Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems, 2*, Article 3. https://doi.org/10.1186/2047-2501-2-3
Reddy, V. M., & Nalla, L. N. (2024). Personalization in e-commerce marketing: Leveraging big data for tailored consumer engagement. *Revista de Inteligencia Artificial en Medicina, 15*(1), 691–725. https://www.researchgate.net/publication/386218629
Rust, R. T., & Huang, M.-H. (2012). Optimizing service productivity. *Journal of Marketing, 76*(2), 47–66. https://doi.org/10.1509/jm.10.0441
Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly, 35*(3), 553–572. https://doi.org/10.2307/23042796
Verhoef, P. C., Kooge, E., & Walk, N. (2015). *Creating value with big data analytics: Making smarter marketing decisions* (1st ed.). Routledge. https://doi.org/10.4324/9781315734750
Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165*, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031
Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing, 80*(6), 97–121. https://doi.org/10.1509/jm.15.0413
Yaiprasert, C., & Hidayanto, A. N. (2023). AI-driven ensemble three machine learning to enhance digital marketing strategies in the food delivery business. *Intelligent Systems with Applications, 18*, Article 200235. https://doi.org/10.1016/j.iswa.2023.200235
Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys, 52*(1), Article 5. https://doi.org/10.1145/3285029