Available Online: 27 October 2026 DOI: 10.54254/2977-5701/2026.38413

A comparative study of linear regression and decision trees for predicting consumer purchase intention across consumption scenarios of different complexity

Luyao Jiang

Kingsway College, Oshawa, Canada

jiangluyao1017@gmail.com

Abstract. With the development of e-commerce, social media marketing and digital operations, enterprises can obtain a large amount of consumer data such as prices, revenues, discounts, product ratings, browsing frequencies and historical purchase records. How to utilize these data to predict consumers' purchase intentions has become an important issue in business analysis and precise marketing. This paper compares the prediction characteristics of linear regression and decision trees in different complexity consumption scenarios. Through literature research, simulation data analysis and comparative analysis, two types of experiments are constructed: the first type only sets an approximately linear relationship between product prices and purchase intentions; the second type adds income, discounts, ratings, browsing frequencies and historical purchase frequencies, and sets threshold effects and variable interactions. The research results show that in simple linear scenarios, linear regression can achieve slightly better prediction performance with lower complexity; in multi-variable, non-linear and interactive relationship scenarios, the prediction error of decision trees is lower. The paper believes that the selection of business models should be based on data structure, prediction accuracy, interpretability and implementation cost, rather than simply pursuing model complexity.

Keywords: consumer purchase intention, linear regression, decision tree, predictive modeling, business analytics

1. Introduction

The development of e-commerce and social media has made the consumer purchasing process increasingly digital. Unlike traditional retail which mainly relies on sales volume, member information, and questionnaires, online platforms can also record behavioral data such as the number of page views, duration of stay, historical purchases, product evaluations, promotional activities, and price changes. These data give enterprises the opportunity to predict consumers' potential willingness before the purchase occurs. However, purchase willingness is usually not determined by a single factor. An increase in price may reduce willingness, but high-income consumers may not be equally sensitive; discounts often have a stimulating effect, but their effectiveness may vary depending on income levels. When both rating and browsing frequency are high, the

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

purchasing tendency of users may further increase. Therefore, real consumption data may exhibit both relatively simple linear trends and thresholds, clustering, and variable interactions.

Data-driven decision-making has become an important part of modern business analysis. Provost and Fawcett pointed out that one of the important values of data science is to convert data into information that can support business decisions [1]. In the e-commerce field, Sakar et al. used page views, sessions, and user information to predict the purchase intentions of online shoppers, demonstrating that behavioral data can provide effective information for purchase prediction [2]. Linear regression and decision trees are relatively easy to understand and implement basic supervised learning methods in business analysis. Linear regression describes the overall average relationship between variables and the target through a mathematical function, having the characteristics of clear structure, low computational cost, and easily interpretable coefficients; decision trees form judgment paths through repeated data division and can handle variable thresholds, consumer clustering, and non-linear interactions [3, 4]. Existing studies have separately explored the connections between feature engineering, classification models, and purchase behavior prediction of consumer psychological variables; these achievements provide a research background for comparing different prediction methods, but cannot directly prove that a certain algorithm is superior in all scenarios [5, 6]. There are significant differences in the prediction logic between the two models, so they are suitable for discussing the question of whether different models are needed for different data complexities.

Based on these differences, this paper mainly discusses three issues: First, in the case of only a single factor and a nearly linear relationship, how much do the predictions of the two models differ; second, when variables increase and non-linear and interaction relationships occur, will decision trees show stronger adaptability; third, if the models are truly used for business analysis, how should the trade-off be made between prediction accuracy, interpretability, and implementation cost. This paper does not directly use real user data from a certain platform, but uses simulation experiments. The main reason for this is that it can control the sample size, variable range, and noise level, temporarily isolating missing values, sample deviations, and other real-world problems, and focusing more on observing the impact of data structure changes on the two models. Therefore, the results of this paper are more suitable for comparing model adaptability, but cannot be directly interpreted as causal relationships in the real market.

2. Theoretical foundations and predictive logic of linear regression and decision trees

2.1. Basic principles of linear regression and its suitability for consumer prediction

Both models predict the same outcome: a continuous purchase-intention score between 0 and 100. I use ordinary linear regression and a regression decision tree so that their predictions can be evaluated on equal terms. The general form of the linear regression model is (Equation (1)):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \tag{1}$$

In this expression, Y stands for the purchase-intention score, β_0 for the intercept, and β_i for the coefficient associated with each explanatory variable X_i. These variables can include price, income, discounts, and browsing frequency. The term ε captures variation not explained by the model. Because the outcome is a continuous score, it makes sense to compare ordinary linear regression directly with a regression tree. If the task were instead to predict whether a customer actually made a purchase (0 or 1), logistic regression would generally provide a more appropriate linear starting point. For the scenarios examined here, the model can be written in its expanded form (Equation (2)):

$$Y = \beta_0 + \beta_1 Price + \beta_2 Income + \beta_3 Discount + \beta_4 Rating + \beta_5 Browsing + \beta_6 PreviousPurchases + \varepsilon \quad (2)$$

The appeal of linear regression is fairly immediate: it gives coefficients that can be discussed in ordinary business terms. A negative Price coefficient means that a higher price is associated with lower predicted intention, with the other included variables held constant. A positive Discount coefficient suggests the reverse association for discounts. Of course, that interpretation is only as useful as the model specification. James et al. place linear regression among the basic methods for examining relationships as well as making predictions [3].

Its limitations are also quite clear. The standard linear regression assumes that each variable affects the outcome in a relatively stable marginal way. If high-income consumers are not sensitive to prices, while low-income consumers significantly increase their purchase intention only when the discount exceeds a certain level, then a single price coefficient or discount coefficient may not accurately represent all consumers. Although the model can be improved by adding squared terms or interaction terms, this requires researchers to identify the corresponding structure beforehand.

2.2. Basic principles of decision trees and their suitability for consumer prediction

The decision tree continuously divides the data by selecting features and thresholds, making the samples at the same node have relatively similar target values. The CART framework is an important foundation for modern regression trees and classification trees [7]. In the prediction of purchase intention, a tree may first determine whether Income is above a certain threshold, then continue to divide based on Rating, Browsing Frequency or Discount, and finally form multiple consumer subgroups.

This structure enables the decision tree to naturally represent conditional relationships. For example, "low income and discount over 20%" may form a group with a significantly increased purchase intention; "high rating and many browsing times" may also form another type of users with high intention. If traditional linear regression is to represent the interaction between Rating and Browsing, it usually needs to actively add the Rating × Browsing term, while the decision tree can automatically form similar rules through continuous splitting.

Another advantage of the decision tree is that the results can be directly converted into judgment paths, which is convenient for business personnel to understand consumer segmentation. However, tree models are also prone to overfitting due to overly deep trees, and the splitting position may change when the sample changes, and the stability is usually lower than simple linear models. Therefore, this paper controls the model complexity by limiting the maximum tree depth and the minimum sample number of leaf nodes. Hastie et al. pointed out that a balance needs to be achieved between the flexibility of the tree model and the control of complexity [4].

2.3. Key differences in how the two models make predictions

The differences between linear regression and decision trees extend beyond the form of the formulas; they also lie in the way they "view" the relationships within the data. Linear regression aims to summarize the average trend with a single overall function, such as how much the purchase intention decreases approximately for every unit increase in price. Decision trees, on the other hand, are more like grouping consumers step by step: first, distinguish between high-income and low-income groups, and then examine factors like ratings, discounts, or browsing frequency separately. These differences imply that the two models may not perform identically under different data structures.

If there are fewer variables and the relationships are relatively stable, linear regression often captures the main trends and the results are easy to interpret. Conversely, when there are obvious thresholds, different consumer groups, or interactions among multiple variables in the data, decision trees' local divisions might be more appropriate. However, this advantage is not unconditional; the depth of the tree, sample size, and noise level all affect the results. Both methods also have a certain degree of interpretability, but the ways of interpretation are different: linear regression focuses on regression coefficients, while decision trees focus on the splitting paths. Rudin's discussion on interpretable models indicates that in scenarios where the decision process needs to be explained, transparency is itself part of the model's value [8]. Recent studies on consumer behavior have also begun to utilize interpretable machine learning to analyze feature importance and differences among consumer groups [9, 10].

2.4. Evaluation metrics and experimental setup

This paper uses the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination R^2 to evaluate the prediction performance of the test set. A smaller MAE indicates a lower average prediction deviation, RMSE imposes a higher penalty for larger prediction errors, and a value of R^2 closer to 1 indicates a higher proportion of target variable changes that the model can explain (Equations (3)-(5)).

$$MAE = (1/n)\Sigma_i|y_i - \hat{y}_i| \tag{3}$$

$$RMSE = \sqrt{[(1/n)\Sigma_i(y_i - \hat{y}_i)^2]} \tag{4}$$

$$R^2 = 1 - [\Sigma_i(y_i - \hat{y}_i)^2/\Sigma_i(y_i - \overline{y})^2] \tag{5}$$

The simulations were carried out in Python using scikit-learn for model fitting, train-test splitting, and evaluation [11]. In both experiments, the data were divided into 80% for training and 20% for testing, with a fixed random seed. One limitation is that the original experiment notes do not record the seed value, complete data-generation code, or all noise settings. This means that the exact numbers in the tables cannot be independently reproduced from the information provided here. The simulations are intended to make the main relationships visible, not to substitute for real customer data. In practice, missing values, measurement errors, selection bias, and unobserved influences could affect the results at the same time. By controlling the main features of the simulated datasets, this paper can compare the models under clearer conditions before considering those additional problems.

3. Comparing predictive performance across consumption scenarios of different complexity

3.1. Simple consumption scenario: prediction with a single linear relationship

The first group of experiments set up a relatively simple retail scenario, only examining the relationship between product prices and purchase intentions. A total of 1,200 observation records were simulated, with Price randomly varying between 20 and 100. Purchase intentions were generated using the formula "95 - 0.72 × Price + random error", and were limited to a range of 0 to 100. This setting was intentionally kept simple: the higher the price, the lower the average purchase intention, but the random error still retained individual differences. This allowed for an initial observation of whether the two models truly required the same level of complexity when the data structure was close to linear.

Linear regression directly estimates the overall downward trend of price and purchase intention. The decision tree was set with a maximum depth of 4 and a minimum leaf node sample size of 25. The main purpose was to limit excessive splitting and avoid the model mistaking random fluctuations as price patterns. Figure 1 shows the fitting situation on the test set, with the vertical axis using a purchase intention score of 0 to 100; Table 1 provides the corresponding MAE, RMSE, and R^2. The tree parameters are used to control the model complexity and were not subjected to large-scale hyperparameter search, so the results reflect the comparison under the settings of this study rather than the possible performance limits of the two algorithms.

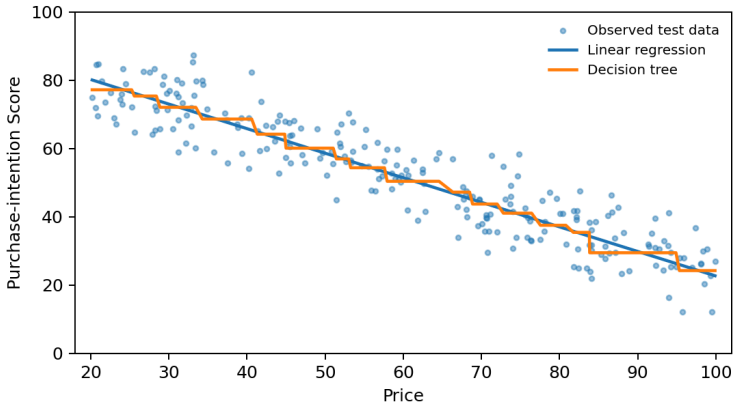


Figure 1. Simple linear consumption scenario and fitted predictions

Table 1. Prediction results in the simple consumption scenario

Model	MAE	RMSE	R^2
Linear Regression	5.399	6.696	0.852
Decision Tree	5.495	6.795	0.848

From Table 1, it can be seen that both models demonstrated a good fit for the negative relationship between price and willingness on this test set. The MAE of the linear regression was 5.399, slightly lower than the 5.495 of the decision tree; the R^2 values were 0.852 and 0.848 respectively. The difference is actually quite small, so we cannot conclude that linear regression will always perform better on all simple data. More accurately, in this simulation setup, the true relationship is close to linear, and linear regression is more compatible with the data generation method, thus achieving a slight advantage.

Figure 1 also shows this difference. The linear regression provides a continuous downward trend, while the decision tree can only approximate it in several price intervals in a stepped manner. For this set of data, this additional segmentation does not provide much new information. This result reminds us that a more flexible model does not necessarily mean better performance on the test set, especially when there are few variables and the structure is relatively stable.

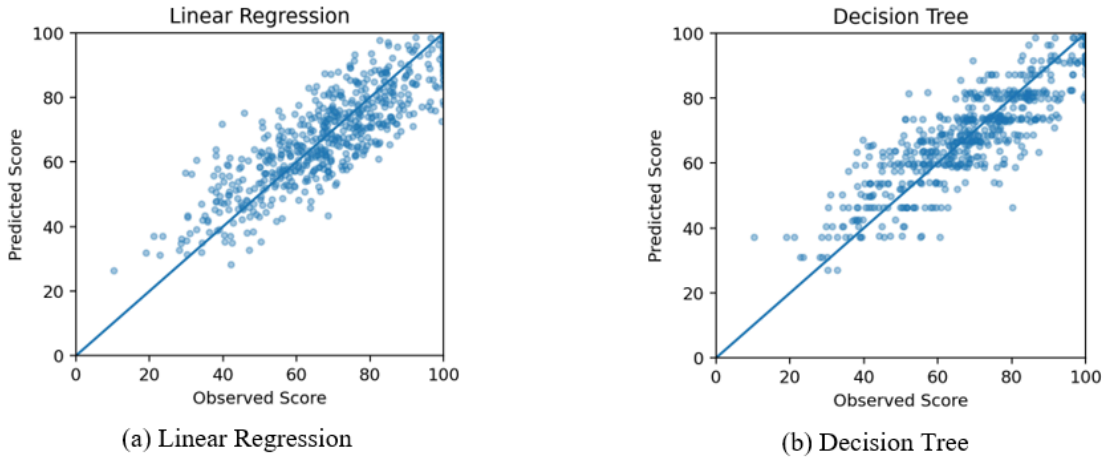
3.2. Complex consumption scenario: prediction with multiple factors and nonlinear interactions

The second group of experiments expanded the scenario to a data structure more similar to that of e-commerce or social media marketing. The sample size was increased to 3,000, and six variables including Price, Income, Discount, Product Rating, Browsing Frequency, and Previous Purchases were added. Unlike the first group, here, instead of merely setting simple linear effects, several thresholds and variable interactions were

artificially introduced to observe what changes would occur when the two models faced more complex relationships.

Specifically, when Income was high, the sensitivity of consumers to Price changes was set lower; when Income was low and Discount reached a certain level, the purchase intention increased additionally; when Product Rating and Browsing Frequency were both high, the intention increased further; users with a higher number of previous purchases and higher ratings also had an additional increment; when Price exceeded the high price threshold, the purchase intention would decrease slightly again. These rules were designed to construct a clear test environment that explicitly includes non-linearity and interaction, and are not a replication of the consumption function of a specific real enterprise. The variable relationships in reality are likely more chaotic and may not follow these thresholds.

Linear regression still uses the six original variables and does not add artificial interaction terms; the decision tree is set with a maximum depth of 7 and a minimum leaf node sample size of 20. This is done to compare: without informing the model of the specific interaction structure in advance, can the two methods adapt to these relationships from the data? Figure 2(a) shows the observed values and predicted values of linear regression, Figure 2(b) shows the results of the decision tree; the closer the points are to the 45-degree reference line, the closer the prediction is to the observation. Table 2 summarizes the corresponding MAE, RMSE, and R^2. Like the first group, the tree parameters here are used to control complexity, not the "optimal parameters" obtained through exhaustive search. Therefore, the subsequent differences should be understood under this premise.



(a) Linear Regression (b) Decision Tree

Figure 2. Observed and predicted purchase-intention scores in the complex scenario

Table 2. Prediction results in the complex consumption scenario

Model	MAE	RMSE	R^2
Linear Regression	7.763	9.609	0.727
Decision Tree	7.114	9.146	0.753

In complex scenarios, the absolute errors of both models are higher than those of the first group; however, the sample sizes, target distributions, and data generation mechanisms of the two groups are different, so the increase in errors cannot be directly attributed to the increase in variables or the enhancement of non-linearity. It is worth noting that the relative performance of the two models has changed: the MAE of the decision tree is 7.114, which is lower than 7.763 of the linear regression; the RMSE is 9.146 and 9.609 respectively; and the

R^2 is 0.753 and 0.727. The decision tree is superior in all three indicators, but the advantage is still limited and not overwhelming.

This result is related to the thresholds and interaction structures added in the experiment. The decision tree can first divide by income and then determine discounts or prices within different income groups, so it is easier to form local rules. The standard linear regression estimates a unified average effect of the original variables and, without manually adding interaction terms, it is difficult to fully represent these conditional relationships. However, this conclusion only applies to the current simulation mechanism; if the interactions in the real data are weak, or if the linear model adds appropriate transformation terms, the gap may narrow.

3.3. Overall comparison of the two models

Table 3. Comprehensive comparison of the two models

Comparison Dimension	Linear Regression	Decision Tree
Simple linear data	Good fit	Can handle it, but extra complexity offers limited benefit
Nonlinear relationships	Requires transformations or nonlinear terms	Can learn through splits
Variable interactions	Usually requires manually added interaction terms	Can capture them through successive splits
Interpretability	Coefficients are straightforward to interpret	Decision paths are straightforward to interpret
Risk of overfitting	Relatively low	Higher when the tree is too deep
Model stability	Relatively high	More sensitive to sample changes
Consumer segmentation	Limited	Strong
Implementation cost	Low	Moderate

Table 3 compares the two models in terms of simple linear data, nonlinear relationships, variable interactions, interpretation methods, overfitting risk, stability, consumer segmentation, and implementation costs. The information provided by the two sets of experiments does not state that "decision trees are more advanced" or "linear regression is more reliable", but rather that the model needs to match the data structure. In the first group, the results of both are almost the same, with linear regression slightly having an advantage; in the second group, after adding thresholds and interactions, the decision tree performs better. That is to say, the complexity of the model itself cannot solely determine the prediction effect; the way data is generated, noise, and parameter settings all play a role.

4. Strengths, limitations, and model selection in business applications

4.1. Practical strengths and limitations of linear regression

One of the direct advantages of linear regression is its ease of interpretation. Enterprises not only obtain predicted values but can also observe the coefficients of each variable, thereby understanding the average relationship between factors such as price, discount, income, and purchase intention. For managers, such results are relatively easy to translate into business language, for example, "When other conditions remain unchanged, if the price increases by one unit, the average change in purchase intention is how much". Of

course, this interpretation relies on linear relationships and model settings. If the consumer group is very diverse, a single coefficient may also mask local differences.

In addition, the computational cost of linear regression is relatively low, and the requirements for sample size and computing resources are relatively limited. When an offline small store only has data on price and sales volume, the core issue is simply to determine whether the price change is related to the demand change. Linear regression can often provide clear trend information at this point. At this time, using complex models may not bring sufficient benefits to offset the implementation costs.

Its limitation lies in its insufficient ability to express complex consumption behaviors. When consumers have obvious group divisions, a unified average coefficient may mask the differences between different groups. For example, high-income consumers are not sensitive to prices, while low-income consumers are highly dependent on promotions. When these two behaviors are combined, only a medium-intensity price coefficient may be obtained. To improve this, it is necessary to artificially add interaction terms such as Income × Price, Rating × Browsing, etc., which will increase the work of model design and interpretation.

4.2. Practical strengths and limitations of decision trees

Decision trees are particularly suitable for handling thresholds and consumer segmentation. For instance, the model might identify a high-intent group such as "high browsing frequency, high rating, and history of purchase". Such a judgment path is quite similar to the customer segmentation approach commonly used by marketers, making it easy to convert into screening rules. However, the thresholds in real projects may not remain stable over time. After a promotion season, changes in platform traffic structure or user composition, the splitting positions of the same tree might change. Therefore, one should not rely solely on the results of a single training session.

Decision trees can also form multiple consumer groups at leaf nodes, such as high-income and low-price-sensitive users, low-income and high-discount-sensitive users, high-browsing and high-rating interested users, and high-history-purchase-frequency users. Enterprises can design differentiated discounts for different nodes instead of applying the same strategy to all consumers.

The flexibility of decision trees also brings issues of overfitting and stability. If the tree keeps getting deeper, each leaf node has very few samples, and the model might learn the accidental fluctuations in the training set. During actual deployment, it is usually necessary to limit the depth of the tree, set a minimum sample size for leaf nodes, or perform pruning, and check the performance using an independent validation set. Another practical problem is that the splitting path of the tree is sensitive to sample changes; even if the overall prediction error does not change much, the specific splitting sequence may change, which will bring additional costs to business interpretation and long-term maintenance.

4.3. Recommendations for selecting models in different business settings

Model selection cannot be solved by a fixed rule. Model selection should first examine the data itself: whether the variables are few, whether the relationships are roughly stable, and whether there are obvious thresholds or consumer segmentation. Then, decisions should be made based on the predictive performance on the validation set, the explanatory needs, and the deployment cost. For data with simpler relationships, linear regression is usually worth being used as the baseline model; if there are strong nonlinearities and conditional combinations in the data, it is more reasonable to consider decision trees next. This sequence also has an actual advantage: using a simple model to establish a baseline can more clearly determine how much additional benefit is brought by increasing the model complexity.

For scenarios with a small amount of data and variables, where the direction of variable effects is relatively stable and the overall relationship is close to linear, such as analyzing the relationship between price changes and sales or purchase intentions for offline stores, linear regression can be tried first. Its advantages are not only low computational cost, but also that the results are easier to check. If the model can already explain most trends, adding more complex methods may not be necessary.

For e-commerce scenarios with multiple variables such as price, page views, ratings, discounts, and historical purchases, if thresholds, consumer segmentation, or interaction relationships do exist in the data, decision trees will be more attractive. It can break down different condition combinations into specific paths, making it easier to identify user groups. However, whether to adopt it ultimately should still depend on the results of the validation set and the stability of the model in different time periods, rather than simply deploying it because the error on the training set is lower.

Two models can also undertake different tasks and do not necessarily have to choose between them. Linear regression is suitable for looking at the overall average relationship, while decision trees can further search for local groups and condition combinations. In practical analysis, this sequence of being simple first and then complex is often easier to explain the results and also convenient for discovering whether the complex model truly provides new information.

The experiments in this paper are based on simulated data, so the thresholds, error values, and model performance rankings set in the paper cannot be directly generalized to real enterprises. Real projects will also encounter missing values, time changes, sample representativeness, selection biases in promotional activities, and feedback effects after the model goes live. For example, distributing coupons to users with high predicted intentions themselves may change subsequent purchase data, affecting the next round of model training. Therefore, this paper aims to illustrate a comparison approach rather than providing fixed commercial decision-making rules.

5. Conclusion

This paper compares the performance of ordinary linear regression and regression-based decision trees on the test set under two pre-defined data generation mechanisms. In the first group, only the approximate linear relationship between price and purchase intention is retained, with a linear regression R^2 of 0.852 and a decision tree R^2 of 0.848, showing a small gap. In the second group, after adding income, discount, rating, browsing frequency, historical purchase frequency, as well as artificially set thresholds and interactions, the R^2 of the decision tree is 0.753, higher than the 0.727 of the linear regression in the same group, and at the same time, MAE and RMSE are also lower. This result supports a relatively limited conclusion: when the data structure is close to the linear assumption, a simple model may be sufficient; when obvious local rules appear in the data, tree models are more likely to adapt. However, this difference is affected by parameter settings and data generation methods and cannot be directly generalized to all consumption prediction tasks.

From the perspective of model selection, complexity should not be the only criterion. A more practical approach is to first establish an interpretable baseline model and then determine whether more complex methods bring stable and repeatable improvements. This paper also has several specific limitations. First, the thresholds and interactions in complex scenarios are set by researchers in advance, which makes the experiment easier to control but also means that the results are partially dependent on these artificial rules; the interaction forms in real consumption behaviors may be completely different. Second, the decision tree parameters are used to control complexity and have not been optimized through cross-validation or systematic hyperparameter search, so the errors in the table do not represent the best performance of the model. Third,

this study uses a continuous "purchase intention score" ranging from 0 to 100, while real platforms are more common in whether to purchase, purchase amount, or repeat purchase and other outcome variables. Different target variables may change the model comparison conclusions. Fourth, this research is mainly based on a fixed data generation and training-test division, and no further repeated simulations are conducted to test the fluctuation of results. In addition, the two groups of experiments not only change the variable relationships but also simultaneously change the sample size, making it impossible to strictly separate the independent influence of "complexity" factor; only comparing the linear model without interaction terms expansion with a single tree with specific parameters does not constitute a fair and comprehensive assessment of the best performance of the two types of models. Moreover, the two groups of experiments simultaneously change the sample size and variable relationships, and cannot independently identify the role of data complexity; only comparing the linear model without interaction terms expansion with a single tree with preset parameters does not suffice to prove the performance upper limit of either type of model. Future research can use real e-commerce data and through repeated sampling, cross-validation, and more model comparisons to verify whether these conclusions are stable.

References

[1] Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data, 1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

[2] Sakar, C. O., Polat, S. O., Katircioglu, M., & Kastro, Y. (2019). Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. *Neural Computing and Applications, 31*(8), 6893–6908. https://doi.org/10.1007/s00521-018-3523-0

[3] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

[4] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

[5] Satu, M. S., & Islam, S. F. (2023). Modeling online customer purchase intention behavior applying different feature engineering and classification techniques. *Discover Artificial Intelligence, 3*, Article 36. https://doi.org/10.1007/s44163-023-00086-0

[6] Azad, M. S., Khan, S. S., Hossain, R., Rahman, R., & Momen, S. (2023). Predictive modeling of consumer purchase behavior on social media: Integrating theory of planned behavior and machine learning for actionable insights. *PLOS ONE, 18*(12), Article e0296336. https://doi.org/10.1371/journal.pone.0296336

[7] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth International Group.

[8] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[9] Chen, Y., Liu, H., Wen, Z., & Lin, W. (2023). How explainable machine learning enhances intelligence in explaining consumer purchase behavior: A random forest model with anchoring effects. *Systems, 11*(6), Article 312. https://doi.org/10.3390/systems11060312

[10] Joung, J., & Kim, H. (2023). Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews. *International Journal of Information Management, 70*, Article 102641. https://doi.org/10.1016/j.ijinfomgt.2023.102641

[11] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.