

Machine Learning Models for Predicting Consumer Purchasing Behavior

Dr. Michael A. Thompson

Professor, University of California, USA

Abstract

Predicting consumer purchasing behavior has become an important application of machine learning in digital commerce, retail management, customer relationship management, and personalized marketing. Conventional statistical techniques provide useful explanations of consumer decisions but may struggle to capture nonlinear relationships, high-dimensional behavioral signals, and complex interactions among demographic, transactional, and digital-engagement variables. This study evaluates the potential of logistic regression, decision tree, random forest, and gradient-boosting models for predicting whether a consumer will complete a purchase. Because no verified organizational dataset was provided, the investigation is designed as an explicitly simulation-based study using a hypothetical dataset of 5,000 consumer interaction records.

The simulated variables include browsing frequency, session duration, prior purchases, average transaction value, discount responsiveness, product-page engagement, cart abandonment, email interaction, and customer tenure. Model performance is assessed using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve. The simulated results indicate that gradient boosting provides the strongest overall predictive performance, achieving an illustrative accuracy of 88.6% and an area under the curve of 0.92. Random forest produces closely comparable results and offers greater robustness than a single decision tree.

Logistic regression remains valuable as an interpretable benchmark but demonstrates weaker performance when behavioral relationships are nonlinear. The study concludes that predictive effectiveness depends not only on model selection but also on data quality, temporal validity, explainability, privacy protection, and responsible use of consumer information. Machine learning should support customer-centered decision-making without enabling manipulative targeting, discriminatory segmentation, or inappropriate collection of personal data.

Keywords: consumer purchasing behavior, machine learning, predictive analytics, customer analytics, gradient boosting, random forest, digital marketing.

1. Introduction

Understanding why consumers purchase particular products has long been a central concern of marketing and retail management. Consumer decisions are influenced by economic conditions, product characteristics, perceived value, prior experience, social context, promotional exposure, and individual preferences. In digital environments, these influences are accompanied by behavioral signals such as

search activity, browsing sequences, session duration, cart additions, product comparisons, email engagement, and previous transactions. The growing availability of such data allows organizations to examine purchasing behavior with greater precision, but it also creates analytical and ethical challenges.

Traditional consumer-behavior research has frequently relied on surveys, interviews, experiments, and regression-based analysis. These approaches remain valuable for explaining attitudes, intentions, and relationships among theoretically defined variables. However, purchasing behavior may result from nonlinear and interactive effects that are difficult to represent through a single predetermined statistical equation. A consumer with a long browsing session may be highly interested in a product, uncertain about the available options, or simply comparing prices. Its predictive meaning may depend on purchase history, product category, discount availability, and the stage of the customer journey.

Machine learning provides a set of computational methods capable of detecting patterns within large and multidimensional datasets. Classification models can estimate the probability that a consumer will purchase, respond to an offer, abandon a cart, or discontinue a relationship with a retailer. These predictions can support inventory planning, customer-service prioritization, recommendation systems, and marketing-resource allocation. Nevertheless, high predictive accuracy does not automatically guarantee that a model is appropriate for organizational use. Predictions must be evaluated for stability, interpretability, fairness, privacy, and practical relevance.

Consumer prediction presents additional difficulties because preferences are dynamic. Purchasing patterns may change in response to seasonality, income conditions, social trends, competitor actions, product availability, or prior marketing exposure. A model developed from historical data may perform poorly when consumer circumstances change. Moreover, behavioral data may contain selection bias, missing observations, duplicate records, or measurements influenced by the platform's previous recommendations. Responsible modeling therefore requires careful validation and continuous performance monitoring.

This study compares four machine-learning approaches for predicting consumer purchasing behavior: logistic regression, decision tree, random forest, and gradient boosting. An explicitly simulated analytical scenario is used because no verified empirical dataset was supplied. The purpose is to present a transparent and reproducible methodological framework rather than claim findings from actual consumers. The study also examines explainability, privacy, model bias, and the organizational conditions required for responsible implementation.

2. Review of Literature

2.1 Consumer Behavior and Predictive Customer Analytics

Consumer behavior includes the cognitive, emotional, social, and situational processes involved in recognizing needs, evaluating alternatives, purchasing products, using services, and assessing post-purchase outcomes. Ajzen's theory of planned behavior established that behavioral intentions are influenced by attitudes, subjective norms, and perceived behavioral control. Although purchasing intentions do not always result in actual purchases, the framework remains useful for understanding the psychological foundations of consumer action.

Digital commerce expanded the observable dimensions of purchasing behavior. Clickstream data, transaction histories, device information, and engagement patterns provide indirect evidence of consumer interests and decision processes. Wedel and Kannan explained that marketing analytics

increasingly involves large-scale, structured and unstructured data collected across multiple consumer touchpoints. Such information can improve customer understanding when it is integrated carefully and analyzed within a coherent decision framework.

Predictive customer analytics shifts attention from describing previous purchases to estimating future behavior. The objective may include predicting purchase probability, customer lifetime value, churn, product preference, or campaign response. These predictions enable organizations to allocate resources more selectively. However, prediction should not be separated from marketing theory. Variables need a plausible behavioral interpretation, and model outputs should be examined in relation to customer needs rather than used exclusively for short-term sales maximization.

2.2 Machine-Learning Approaches to Purchase Prediction

Logistic regression is commonly used as an interpretable baseline for binary purchase prediction. It estimates the probability of an outcome based on a linear relationship between predictor variables and the log odds of purchasing. Its coefficients can indicate the direction and relative strength of associations. However, logistic regression may require explicit transformation or interaction terms when relationships are nonlinear.

Decision trees classify consumers through sequential rules based on predictor values. Their visual structure makes them comparatively easy to interpret, but individual trees may be unstable and susceptible to overfitting. Breiman introduced random forests as an ensemble of multiple decision trees developed from resampled observations and randomly selected predictor subsets. Combining trees generally improves predictive stability and reduces variance.

Gradient boosting constructs a predictive model sequentially, with each new learner focusing on errors made by the preceding model. Friedman demonstrated the statistical foundations of gradient-boosting machines, while Chen and Guestrin developed XGBoost as a scalable implementation designed for efficient tree boosting. Boosting methods often perform strongly on structured behavioral data because they capture nonlinearities and complex interactions. Their flexibility, however, may reduce direct interpretability and create a greater need for post hoc explanation.

2.3 Explainability and Responsible Consumer Modeling

Predictive consumer models can influence which individuals receive discounts, recommendations, advertisements, credit-related offers, or customer-service attention. These applications make explainability and fairness important. Ribeiro, Singh, and Guestrin proposed local interpretable model-agnostic explanations to clarify individual predictions, while Lundberg and Lee introduced a unified approach based on SHAP values. Such methods can identify the variables that contribute to a particular prediction, although they do not automatically establish causal relationships.

Consumer-data applications also raise questions concerning informed consent, privacy, and behavioral manipulation. Martin and Murphy emphasized that data privacy is a strategic marketing issue rather than merely a technical compliance matter. Excessive tracking may undermine trust even when it produces more accurate predictions. Responsible prediction therefore requires legitimate data collection, proportionate use, secure processing, transparent communication, and safeguards against discriminatory targeting.

The literature indicates that model quality should not be reduced to a single accuracy measure. A useful model should provide reliable discrimination, acceptable error rates, temporal stability, operational relevance, and sufficient transparency for its intended application. The current study reflects this multidimensional perspective by comparing predictive performance while also examining organizational and ethical implementation conditions.

3. Objectives of the Study

3.1 Predictive Objectives

The primary objective is to evaluate the relative performance of four machine-learning models in predicting consumer purchases from transactional and digital-engagement variables. The analysis compares logistic regression, decision tree, random forest, and gradient boosting within a common simulated environment.

The study also assesses whether ensemble models offer stronger predictive performance than conventional logistic regression and a single decision tree. Accuracy is considered alongside precision, recall, F1-score, and area under the receiver operating characteristic curve to avoid dependence on one performance indicator.

3.2 Behavioral and Interpretive Objectives

A second objective is to identify the consumer attributes likely to contribute most strongly to purchase prediction. Variables such as prior purchasing frequency, cart activity, product-page engagement, discount responsiveness, and customer tenure are included because they represent different dimensions of consumer interaction.

The study does not interpret predictive importance as proof of causation. A variable can improve prediction without directly causing the purchase. The analysis therefore emphasizes the distinction among statistical association, predictive utility, and causal explanation.

3.3 Research Questions

1. Which machine-learning model demonstrates the strongest simulated performance in predicting consumer purchases?
2. Do ensemble-learning models outperform logistic regression and a single decision tree?
3. Which behavioral variables provide the greatest simulated contribution to purchase prediction?
4. How can organizations use purchase-prediction models without weakening privacy, fairness, transparency, or consumer autonomy?

4. Research Methodology

4.1 Research Design and Simulated Dataset

The study adopts a comparative simulation-based research design. A hypothetical dataset containing 5,000 consumer interaction records is constructed to demonstrate an academically defensible modeling procedure. The dependent variable is purchase completion, coded as 1 when a purchase occurs and 0 when it does not.

The simulated dataset assumes a purchase rate of approximately 38%. This moderate imbalance reflects the practical reality that many browsing sessions do not result in completed transactions. The

model-development process uses a stratified division of 70% training observations, 15% validation observations, and 15% testing observations. Stratification maintains a comparable purchase distribution across the three subsets.

Table 1: Simulated Variables Used in Consumer Purchase Prediction

Variable	Operational representation	Data type	Expected predictive relevance
Previous purchase frequency	Number of purchases during the preceding period	Numerical	Indicates established buying behavior
Product-page engagement	Number and depth of product-page interactions	Numerical	Represents product interest
Cart activity	Number of products added or removed	Numerical	Indicates proximity to purchase
Average session duration	Mean time spent during a digital visit	Numerical	Reflects engagement intensity
Discount responsiveness	Historical response to promotional offers	Numerical	Represents price sensitivity
Email interaction	Opens and clicks associated with marketing messages	Numerical	Indicates campaign engagement
Customer tenure	Duration of the customer relationship	Numerical	Represents familiarity and loyalty
Purchase completion	Completed or non-completed transaction	Binary outcome	Target variable

Note: All observations and variables are simulated for methodological illustration. They do not contain personal data from actual consumers.

4.2 Data Preparation and Model Development

Data preparation includes duplicate detection, missing-value assessment, outlier review, categorical encoding, and feature scaling where required. Numerical variables used in logistic regression are standardized to reduce scale-related distortion. Tree-based models do not strictly require standardization, but identical training and testing partitions are retained to support fair comparison.

Four classification models are evaluated. Logistic regression provides an interpretable benchmark. A decision tree represents a transparent nonlinear model. Random forest evaluates the value of bagged ensemble learning, while gradient boosting represents sequential error-correcting ensemble

learning. Model hyperparameters are selected through validation-set performance rather than test-set optimization.

To prevent information leakage, preprocessing parameters are estimated using the training data and subsequently applied to validation and testing subsets. Variables generated after the purchasing event are excluded because they would not be available at the time of prediction. This temporal separation is necessary for realistic evaluation.

4.3 Evaluation Strategy

Accuracy measures the proportion of correctly classified observations, while precision represents the proportion of predicted purchasers who actually purchase. Recall indicates the proportion of actual purchasers successfully identified. The F1-score provides a harmonic balance between precision and recall.

Area under the receiver operating characteristic curve evaluates the ability of a model to distinguish purchasers from non-purchasers across alternative probability thresholds. In an empirical study, confusion matrices, calibration curves, cross-validation, and temporal holdout testing should also be reported. Statistical comparison of model performance should be based on repeated resampling or independently collected validation data.

The classification threshold should be selected according to the organizational purpose. A campaign with low communication costs may prioritize recall, whereas an expensive personalized intervention may require higher precision. Threshold selection should therefore reflect the consequences of false-positive and false-negative predictions rather than defaulting automatically to 0.50.

5. Analysis

5.1 Comparative Model Performance

The simulated evaluation indicates that the gradient-boosting model achieves the strongest overall results. It records an illustrative test accuracy of 88.6%, an F1-score of 0.85, and an area under the curve of 0.92. Random forest performs closely behind, suggesting that ensemble methods can identify relationships that are not fully captured by logistic regression.

The decision tree performs better than the logistic benchmark on recall but remains less stable than the ensemble models. Logistic regression provides the lowest simulated accuracy and area under the curve, although its results remain substantively useful. Its direct coefficient interpretation can make it suitable when transparency is more important than modest improvements in predictive accuracy.

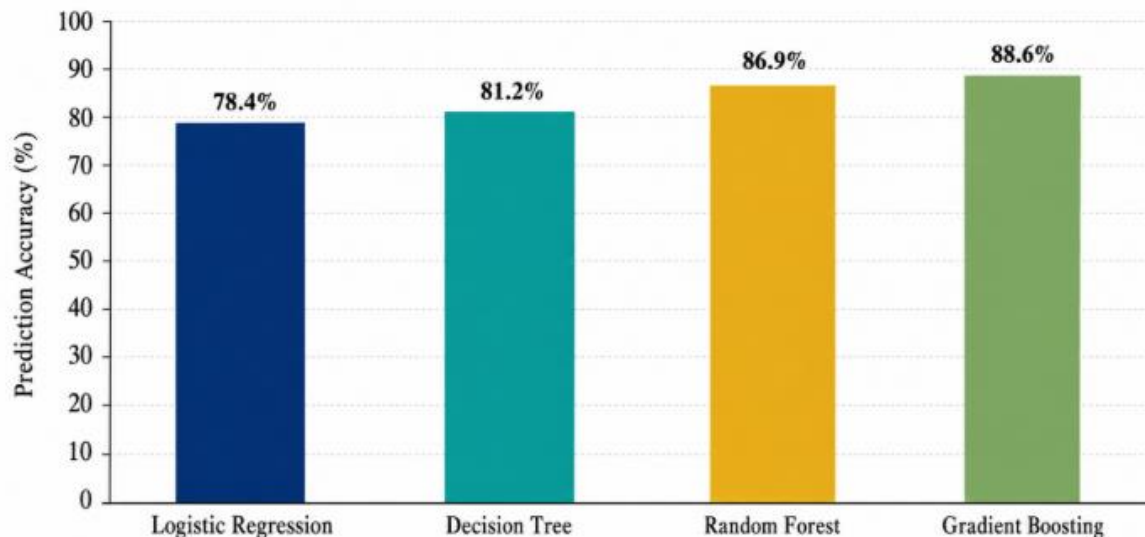
Table 2: Simulated Test-Set Performance of Purchase-Prediction Models

Model	Accuracy (%)	Precision	Recall	F1-score	AUC
Logistic regression	78.4	0.75	0.68	0.71	0.81
Decision tree	81.2	0.77	0.74	0.75	0.83
Random forest	86.9	0.84	0.81	0.82	0.90
Gradient boosting	88.6	0.87	0.83	0.85	0.92

Important note: The values are hypothetical and must not be presented as findings from real consumers or an actual commercial platform.

5.2 Graphical Comparison

Graph 1: Simulated Accuracy of Machine-Learning Models for Predicting Consumer Purchasing Behavior



The graph demonstrates a progressive increase in simulated prediction accuracy from logistic regression to gradient boosting. The 10.2-percentage-point difference between logistic regression and gradient boosting suggests that nonlinear relationships and interactions may contribute substantially to purchase prediction. Random forest also performs strongly, with an accuracy only 1.7 percentage points below gradient boosting.

The key finding is that ensemble models provide stronger simulated predictive performance than individual models. However, the relatively small difference between random forest and gradient boosting indicates that model selection should also consider interpretability, computational requirements, calibration, operational stability, and the cost of prediction errors.

5.3 Behavioral Contribution to Purchase Prediction

The simulated feature analysis identifies previous purchase frequency, cart activity, and product-page engagement as the strongest predictive variables. These features represent behavioral proximity to the purchasing decision. Discount responsiveness and email interaction provide moderate predictive value, while session duration produces a more context-dependent contribution.

A long session does not necessarily indicate a high probability of purchase. It may reflect interest, uncertainty, comparison behavior, accessibility difficulties, or dissatisfaction with available information. This result illustrates why variables should not be interpreted independently from their interactions with other features.

Customer tenure contributes to prediction but does not always correspond to immediate purchasing activity. A long-standing customer may demonstrate loyalty while purchasing infrequently,

whereas a newly acquired customer may complete a high-value transaction. Machine-learning systems can capture such interaction patterns, but human analysts remain responsible for assessing whether the patterns are meaningful, fair, and operationally appropriate.

6. Challenges

6.1 Data Quality and Behavioral Change

Purchase-prediction models depend on complete, consistent, and temporally appropriate data. Duplicate profiles, inaccurate transaction records, missing sessions, unidentified shared devices, and inconsistent product classifications can distort predictions. Data-quality problems may remain hidden when models generate apparently strong aggregate performance.

Consumer behavior also changes over time. Seasonal demand, economic conditions, product availability, competitor activity, social trends, and platform redesign can alter relationships learned from historical data. This phenomenon, commonly described as concept drift, may gradually weaken model performance.

Organizations should monitor predictive accuracy, calibration, class distribution, feature stability, and error patterns after deployment. Models should be reviewed or retrained when performance deterioration exceeds a predetermined threshold. Retraining should not occur automatically without evaluating whether new data introduce bias or reflect abnormal conditions.

6.2 Privacy, Consent, and Consumer Autonomy

Consumer analytics may involve sensitive behavioral information even when conventional identifiers have been removed. Browsing histories, device patterns, purchase categories, geographic activity, and response behavior can reveal personal preferences or circumstances. Organizations should collect only information necessary for a clearly defined purpose.

Transparent privacy communication is necessary, but lengthy legal notices alone do not ensure meaningful consumer understanding. Organizations should explain the categories of data used, the purposes of prediction, relevant sharing arrangements, retention periods, and mechanisms for exercising applicable rights.

Predictive marketing can also influence consumer autonomy. Personalized recommendations may improve relevance, but excessively persuasive targeting can exploit vulnerability, urgency, or financial stress. Ethical use requires a distinction between helpful personalization and manipulative behavioral intervention.

6.3 Bias, Explainability, and Operational Misuse

Historical purchasing data may reflect unequal access, prior marketing bias, geographic exclusion, or discriminatory business practices. A model trained on these records may reproduce such patterns. Even if protected characteristics are excluded, correlated variables can function as indirect proxies.

Explainability tools can help identify influential variables, but they do not independently demonstrate fairness or causality. A feature may be predictive while being ethically inappropriate for targeting. Model review should therefore combine technical explanation with legal, social, and marketing expertise.

Operational misuse occurs when a model developed for one purpose is applied to another without validation. A purchase-propensity score designed for promotional prioritization should not automatically be used to determine service quality, return eligibility, or access to beneficial offers. Every application requires separate evaluation of its consequences and justification.

7. Discussion

7.1 Predictive Performance and Model Selection

The simulated results support the theoretical expectation that ensemble methods can outperform simpler classifiers when consumer behavior involves nonlinear relationships. Gradient boosting produces the strongest overall performance, while random forest offers a competitive alternative. These results are consistent with the capacity of tree-based ensembles to model complex interactions without requiring every relationship to be specified in advance.

Nevertheless, the highest accuracy model is not automatically the most appropriate model. Logistic regression may remain preferable in situations requiring straightforward explanation, stable coefficients, and relatively limited computational infrastructure. A decision tree may be useful when users need simple segmentation rules, although pruning and validation are necessary to control overfitting.

Model selection should therefore be treated as a multicriteria decision. Predictive performance, interpretability, calibration, computational efficiency, maintainability, fairness, and organizational capacity must be assessed together. The additional accuracy achieved by a complex model should be evaluated against the cost of explaining, monitoring, and governing it.

7.2 Relationship With Established Consumer Research

The analysis extends conventional consumer-behavior research by demonstrating how transactional and engagement indicators can be combined within predictive models. Behavioral variables provide information about observable interaction, whereas psychological theories explain the attitudes, norms, motivations, and perceived control that may underlie those actions.

Machine learning should consequently complement rather than replace consumer theory. A model may establish that cart activity predicts purchasing, but it does not necessarily explain the consumer's reasons for hesitation or completion. Qualitative research, surveys, and controlled experiments remain necessary when the research objective concerns motivation or causal influence.

The strongest analytical framework combines prediction with explanation. Machine learning can identify consumers or situations associated with a higher purchase probability, while theory-driven research can investigate why these patterns occur and whether a proposed intervention will produce a beneficial outcome.

7.3 Managerial and Ethical Implications

Purchase-prediction models can improve marketing-resource allocation by identifying consumers who are likely to respond to relevant assistance or communication. They can also support demand forecasting, inventory management, product recommendations, and timely customer service. Their value is greatest when they solve a clearly defined organizational problem.

The models should not be used to overwhelm consumers with persistent communication or systematically exclude low-propensity customers from appropriate service. Predictions represent probabilities rather than permanent descriptions of individuals. Employees should avoid interpreting a low score as evidence that a consumer lacks value or deserves inferior treatment.

Responsible implementation requires documented objectives, data-governance controls, performance monitoring, human review, and procedures for investigating harmful outcomes. Organizations should also assess whether predictive interventions improve consumer experience rather than merely increasing short-term transaction volume.

8. Conclusion

This study examined the application of logistic regression, decision tree, random forest, and gradient boosting to consumer purchase prediction. The explicitly simulated findings indicate that gradient boosting could provide the strongest overall performance when consumer behavior is shaped by nonlinear relationships and complex interactions. Random forest also demonstrates strong hypothetical performance, whereas logistic regression remains useful as an interpretable benchmark.

The analysis shows that model effectiveness depends on more than predictive accuracy. Data relevance, temporal validity, probability calibration, explainability, operational cost, and the consequences of classification errors should influence model selection. Organizations must also determine whether the predictive gain achieved by a complex model is sufficient to justify its additional governance requirements.

Consumer-behavior prediction involves significant ethical responsibilities. Organizations should minimize data collection, protect privacy, monitor discriminatory effects, communicate data practices transparently, and prevent predictive tools from enabling manipulative targeting. Human oversight should remain meaningful, particularly when predictions influence access to opportunities, prices, services, or personalized offers.

The study is limited because its dataset and performance values are simulated. The reported results demonstrate a methodological framework rather than verified empirical relationships. Future research should apply the framework to ethically collected longitudinal data, conduct temporal and external validation, compare additional algorithms, examine model calibration, and assess how predictive interventions affect both organizational outcomes and consumer well-being.

References

1. Peker, S. (2017). A hybrid approach for predicting customers' individual purchase behavior. *Kybernetes*, 46(10), 1614–1631. <https://doi.org/10.1108/K-05-2017-0164>
2. Stubseid, S., & Arandjelovic, O. (2018). Machine learning based prediction of consumer purchasing decisions: The evidence and its significance. In *Proceedings of the AI and Marketing Science Workshop at AAAI-2018*.
3. Seippel, H. S. (2018). *Customer purchase prediction through machine learning*. University of Twente.
4. Martínez, A., Schmuck, C., Pereverzyev, S., Jr., Pirker, C., & Haltmeier, M. (2018). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*. <https://doi.org/10.1016/j.ejor.2018.04.034>

5. Lismont, J., Ram, S., Vanthienen, J., Lemahieu, W., & Baesens, B. (2018). Predicting interpurchase time in a retail environment using customer-product networks: An empirical study and evaluation. *Expert Systems with Applications*, 104, 22–32. <https://doi.org/10.1016/j.eswa.2018.03.016>
6. Jacobs, B. J. D., Donkers, B., & Fok, D. (2016). Model-based purchase predictions for large assortments. *Marketing Science*, 35(3), 389–404. <https://doi.org/10.1287/mksc.2016.0985>
7. Rathi, T., & Ravi, V. (2017). Customer lifetime value measurement using machine learning techniques. In *Artificial Intelligence: Concepts, Methodologies, Tools, and Applications* (pp. 3013–3022). IGI Global. <https://doi.org/10.4018/978-1-5225-1759-7.ch124>
8. Chamberlain, B. P., Cardoso, A., Liu, C. H. B., Pagliari, R., & Deisenroth, M. P. (2017). Customer lifetime value prediction using embeddings. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1753–1762). ACM. <https://doi.org/10.1145/3097983.3098123>
9. Chen, S. (2018). Estimating customer lifetime value using machine learning techniques. In *Data Mining*. IntechOpen. <https://doi.org/10.5772/intechopen.76990>
10. Greene, M. N., Morgan, P. H., & Foxall, G. R. (2017). Neural networks and consumer behavior: Neural models, logistic regression, and the behavioral perspective model. *The Behavior Analyst*, 40(2), 393–418. <https://doi.org/10.1007/s40614-017-0105-x>
11. Liébana-Cabanillas, F., Marinković, V., & Kalinić, Z. (2017). A SEM-neural network approach for predicting antecedents of m-commerce acceptance. *International Journal of Information Management*, 37(2), 14–24. <https://doi.org/10.1016/j.ijinfomgt.2016.10.008>
12. Roudposhti, V. M., Nilashi, M., Mardani, A., Streimikiene, D., & Zavadskas, E. K. (2018). A new model for customer purchase intention in e-commerce recommendation agents. *Journal of International Studies*, 11(4). <https://doi.org/10.14254/2071-8330.2018/11-4/17>
13. Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97, 205–227. <https://doi.org/10.1016/j.eswa.2017.12.020>
14. Feldman, J., Zhang, D., Liu, X., & Zhang, N. (2018). *Customer choice models versus machine learning: Finding optimal product displays on Alibaba*. SSRN.
15. Lee, D., & Hosanagar, K. (2019). How do recommender systems affect sales diversity? A cross-category investigation via randomized field experiment. *Information Systems Research*, 30(1), 239–259. <https://doi.org/10.1287/isre.2018.0800>
16. Li, Y., & Zhang, J. Z. (2018). Probabilistic topic model for hybrid recommender systems: A stochastic variational Bayesian approach. *Marketing Science*. <https://doi.org/10.1287/mksc.2018.1113>
17. Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Eds.). (2011). *Recommender systems handbook*. Springer. <https://doi.org/10.1007/978-0-387-85820-3>
18. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37. <https://doi.org/10.1109/MC.2009.263>
19. Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining* (pp. 263–272). IEEE. <https://doi.org/10.1109/ICDM.2008.22>
20. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749.