



Application of The Random Forest Algorithm in Classifying the Tendency of Impulsive Purchasing Behavior Among Gen Z Consumers in E-Commerce Based on Flash Sale Features

Syifa Nurfadhilah¹, Mulkan Azhari²

^{1,2} Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Article Info

Article history

Received : Apr 18, 2026

Revised : Apr 28, 2026

Accepted : Apr 30, 2026

Keywords:

Random Forest ;
Impulsive Purchases;
Gen Z;
E-Commerce;
Flash Sale.

Abstract

The rapid growth of e-commerce in Indonesia, particularly on Shopee, has significantly influenced consumer behavior through promotional strategies such as flash sales. This study aims to classify impulsive buying tendencies among Generation Z, identify key influencing factors, and develop a web-based classification system for behavioral analysis. A quantitative data mining approach was applied using the Random Forest algorithm. The dataset consisted of 420 Gen Z respondents collected through a Likert-scale questionnaire using purposive sampling, and model evaluation was conducted using 10-fold cross-validation to ensure reliability. The results show that the Random Forest model achieved an accuracy of 83.16%, outperforming Decision Tree (78.42%) and Logistic Regression (75.08%), indicating its effectiveness in handling complex behavioral patterns. Feature importance analysis revealed that limited stock availability (39.85%) and discount magnitude (33.21%) are the most dominant factors influencing impulsive buying behavior, followed by promotional duration and notification attractiveness. These findings emphasize the role of urgency and scarcity in driving impulsive purchases among Gen Z consumers. Additionally, a web-based system was developed using the Flask framework in Python to support automated data processing, model training, and visualization of results. The system enables real-time behavioral analysis and decision support for digital marketing strategies. Overall, the study demonstrates that machine learning, particularly Random Forest, provides a more accurate and objective approach for analyzing impulsive buying behavior compared to conventional statistical methods, while also offering a practical tool for e-commerce analytics and strategy optimization.

Corresponding Author:

Syifa Nurfadhilah,
Program Studi Sistem Informasi,
Universitas Muhammadiyah Sumatera Utara,
Jl. Kapt. Mukhtar Basri No. 3, Glugur Darat II, Kec. Medan Timur, Kota Medan, Sumatera Utara, 20238, Indonesia
Email: syifanur582@gmail.com

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



1. Introduction

Advances in information and communication technology have significantly transformed consumer purchasing behavior, particularly through the rapid expansion of e-commerce platforms.

E-commerce enables consumers to purchase goods and services without spatial and temporal limitations, thereby increasing efficiency, convenience, and accessibility in transaction processes (Hajli, 2015; Laudon & Traver, 2021). In Indonesia, the growth of e-commerce continues to accelerate in parallel with increasing internet penetration and digital device usage, especially among younger demographics (Cahayani & Isyanto, 2025; Statista, 2024).

Among various e-commerce platforms, Shopee has become one of the most dominant players in Indonesia's digital marketplace ecosystem. Shopee actively implements aggressive digital marketing strategies such as free shipping, cashback, discount vouchers, and flash sale campaigns to attract consumer attention and increase conversion rates (Ramadhan & Wardi, 2025; Kotler et al., 2021). Flash sales, in particular, have been recognized as one of the most effective promotional mechanisms due to their ability to create urgency and scarcity effects, which strongly influence consumer decision-making (Lynn, 2015; Zhang et al., 2018).

A flash sale is defined as a short-term promotional strategy offering significant discounts within a limited time frame and quantity. This mechanism triggers psychological pressure through scarcity cues, time constraints, and competitive purchasing environments (Aggarwal et al., 2011; Inman et al., 2014). Although flash sales are effective in increasing transaction volume, they also contribute to impulsive buying behavior among consumers.

Impulsive buying behavior refers to spontaneous, unplanned purchasing decisions driven primarily by emotional and situational stimuli rather than rational evaluation (Rook, 1987; Beatty & Ferrell, 1998). In e-commerce environments, impulsive buying is frequently triggered by visual cues, promotional notifications, countdown timers, and platform-driven gamification mechanisms (Chan et al., 2017; Zhang et al., 2018). Such behavior is increasingly prevalent among digital consumers who are constantly exposed to real-time marketing stimuli.

Generation Z (Gen Z) is identified as the demographic most vulnerable to impulsive buying behavior due to their high digital engagement and strong exposure to social media-driven consumer culture (Priporas et al., 2017; Djafarova & Bowes, 2021). As digital natives, Gen Z consumers tend to exhibit higher responsiveness to promotional stimuli, particularly in mobile-based e-commerce environments such as Shopee. Their purchasing decisions are often influenced by hedonic motivation, peer influence, and instant gratification tendencies (Wang & Zhang, 2022).

Despite the growing body of literature on impulsive buying behavior in e-commerce, most prior studies rely on traditional statistical approaches such as regression analysis and structural equation modeling (Verplanken & Herabadi, 2001; Mai et al., 2003). These approaches often assume linear relationships and are limited in capturing complex, non-linear behavioral patterns. Furthermore, recent empirical comparisons between machine learning models show that ensemble-based methods such as Random Forest outperform single classifiers (e.g., Decision Tree, Logistic Regression, and Naïve Bayes) in behavioral classification tasks due to higher robustness and better generalization ability (Breiman, 2001; Biau & Scornet, 2016; Probst et al., 2019; Zeng et al., 2022). However, studies specifically applying Random Forest to impulsive buying classification in e-commerce environments remain limited, particularly those utilizing Gen Z behavioral data from Shopee.

This gap is critical because Gen Z-specific purchasing behavior in Shopee represents a highly dynamic, high-frequency, and stimulus-rich dataset that differs significantly from traditional consumer datasets. The integration of platform-based behavioral data with machine learning classification provides a more objective and scalable approach for analyzing impulsive buying tendencies. Therefore, this study addresses this gap by applying the Random Forest algorithm to classify Gen Z impulsive purchasing behavior based on flash sale features in Shopee.

In addition to the methodological contribution, this study introduces a novel web-based intelligent classification system that automatically predicts impulsive buying tendencies in real time. Unlike previous studies that focus solely on offline analysis or static modeling, the proposed system provides an interactive decision-support tool that can be implemented for behavioral analytics in e-

commerce environments. This represents a practical contribution to digital marketing analytics and consumer behavior monitoring systems.

The Random Forest algorithm is selected due to its strong performance in handling high-dimensional and noisy data, its ability to model complex interactions, and its reduced risk of overfitting through ensemble learning (Breiman, 2001; Liaw & Wiener, 2002). Its effectiveness has been demonstrated in various classification problems involving consumer behavior and digital analytics (Probst et al., 2019).

The objectives of this study are:

1. To classify Gen Z's tendency toward impulsive purchasing behavior based on flash sale features in Shopee.
2. To identify the most influential flash sale factors contributing to impulsive buying behavior.
3. To develop a web-based system for automatic classification of impulsive purchasing behavior using the Random Forest algorithm.

2. Research Methodology

2.1 Type of Research

This study employs a quantitative method using a data mining approach. The quantitative approach was chosen because the study focuses on processing numerical data obtained from respondents' questionnaires (Abdi et al., 2024). This method was selected because it produces measurable data, allowing the analysis process to be conducted objectively and systematically. Data analysis in this study was conducted using the Random Forest machine learning algorithm, which was used to classify Gen Z's tendencies toward impulsive purchasing behavior based on the "flash sale" feature on Shopee.

The research stages consisted of data collection, data preprocessing, data transformation, classification modeling, and evaluation of classification results. During the data collection stage, the researcher distributed an online questionnaire to respondents who met the study criteria. Subsequently, the collected data was processed through a preprocessing stage to ensure data quality before being used in the classification process. The final result of the study is a website-based impulsive purchasing behavior classification system capable of assisting in the automatic and objective analysis of consumer behavior.

2.2 Population and Sample

The population in this study consists of Gen Z individuals who actively use Shopee and have participated in flash sale programs. In this study, Gen Z refers to individuals aged 13–28 years. The selection of Gen Z as the research subject is based on the characteristics of this generation, which is highly engaged with digital technology, social media, and online shopping activities, making them more prone to impulsive purchasing behavior. The sample size was determined using the Slovin formula with a margin of error of 10%.

$$n = \frac{N}{1 + N(e)^2}$$

Based on our calculations, the minimum sample size required is approximately 100 respondents. To improve the quality of the research data and account for invalid data, the number of respondents used ranges from 100 to 200.

Respondent inclusion criteria include:

1. Belonging to the Gen Z category.
2. Having an active Shopee account.
3. Having made a purchase through the flash sale feature.
4. Willing to complete the questionnaire in full.

Meanwhile, the exclusion criteria include:

1. Not a Shopee user.
2. Has never participated in a flash sale.
3. Filled out the questionnaire incompletely.

These criteria were established to ensure that the data collected aligns precisely with the research focus and accurately reflects Gen Z's impulsive purchasing behavior on the Shopee platform.

2.3 Data Collection Methods

Primary data was collected through an online questionnaire using Google Forms. Google Forms was chosen because it is more practical, easily accessible, and aligns with the characteristics of Gen Z respondents, who actively use digital devices and the internet (Ainurrohmah, C, 2024). Additionally, the online method allows for faster and more efficient data collection.

The questionnaire used a 1–5 Likert scale with the following categories:

1. Strongly Agree = 5
2. Agree = 4
3. Neutral = 3
4. Disagree = 2
5. Strongly Disagree = 1

The Likert scale is used to measure respondents' level of agreement with statements related to impulsive buying behavior and the influence of flash sale features. The research variables consist of independent and dependent variables. The independent variables are flash sale features, which include:

1. Discount (X₁)
2. Duration (X₂)
3. Visuals/Notifications (X₃)
4. Stock Availability (X₄)

Each variable has indicators related to promotional appeal, time pressure, visual presentation, and fear of missing out. These variables were selected because they are considered capable of influencing impulsive purchasing decisions among e-commerce consumers.

The dependent variable in this study is impulsive buying behavior (Y), which is measured based on the indicators of spontaneous purchasing, emotional decision-making, low self-control, immediate gratification, and post-purchase regret. All respondent answers were then converted into numerical data used in the classification process using the Random Forest algorithm.

Table 2.1 Research Variables

Variable	Code	Indicator
Discount	X ₁	Interest in discounts, perception of low prices
Time Duration	X ₂	Time pressure, countdown timer
Visuals/Notifications	X ₃	Promotional notifications, visual displays
Stock Availability	X ₄	Information about limited stock, fear of running out
Impulsive Buying Behavior	Y	Spontaneous purchases, emotional decisions.

2.4 Data Preprocessing

Before the classification process begins, the data undergoes a preprocessing stage aimed at improving data quality to ensure more accurate classification results. The preprocessing stage includes:

1. Data Cleaning

This stage involves removing empty data, invalid data, and inconsistent responses. The data cleaning process is crucial because data quality directly impacts the performance of the classification model used.

2. Data Transformation

Responses to the Likert-scale questionnaire are converted into numerical data so they can be processed using machine learning algorithms. Data transformation is performed to adapt the data format to the needs of the classification system.

3. Data Labeling

Respondents are categorized into impulsive and non-impulsive classes based on their total response scores. The labeling process is performed to define the classification targets that the Random Forest algorithm will learn.

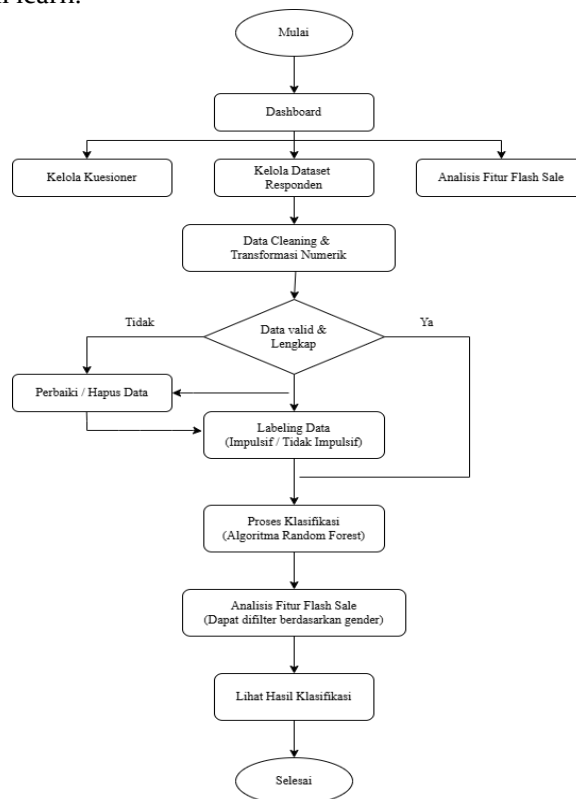


Figure 1: Process Flowchart for Classifying Impulsive Purchasing Behavior Using Random Forest

Figure 1 illustrates the stages of the classification process for impulsive purchasing behavior in the developed system. The process begins with the management of questionnaire data and respondent datasets, followed by data cleaning and numerical transformation to ensure the data is valid and ready for use. The data is then labeled as “impulsive” or “non-impulsive” before being processed using the Random Forest algorithm. The classification results are then presented in the form of an analysis of impulsive behavior based on flash sale features.

2.5 The Random Forest Algorithm

Random Forest is a machine learning algorithm based on ensemble learning that combines multiple decision trees to produce more stable and accurate classifications (Dewi & Avianto, 2024). Each decision tree makes predictions independently, and the final result is determined by majority voting. This approach gives Random Forest better classification capabilities than using a single decision tree.

The advantages of the Random Forest algorithm include:

1. High accuracy.
2. Ability to handle complex data.
3. Reduced risk of overfitting.
4. Ability to identify feature importance.

In this study, Random Forest was used to classify tendencies toward impulsive purchasing behavior based on respondents’ answer patterns regarding the “flash sale” feature. This algorithm was chosen because it can process consumer behavior data that involves many variables and complex relationships. Additionally, Random Forest can demonstrate the level of influence each variable has

on the classification results through feature importance. Thus, researchers can identify which flash sale factors most influence Gen Z's impulsive behavior.

In this study, Random Forest was implemented using the Python programming language with the assistance of machine learning libraries for model training and evaluation. The system was developed to automate and streamline the analysis of impulsive purchasing behavior. Through this system, users can process data and view classification results more quickly and efficiently. The web system is designed with a simple interface to make it easy to use for both users and administrators. Additionally, the system facilitates the visualization of classification results, making the generated information easier to understand. The system was developed using the Python programming language and integrated with the Random Forest algorithm to support the automated classification of impulsive purchasing behavior.

3. Results and Discussion

3.1 Respondent Characteristics

The study's respondents consisted of Gen Z individuals who actively use Shopee and frequently participate in flash sales. Most respondents were students who are accustomed to using digital platforms in their daily activities. The analysis results indicate that the majority of respondents are interested in participating in flash sales due to the significant discounts and the limited duration of the promotions. This indicates that Gen Z is highly responsive to urgency-based digital marketing strategies (Wangi & S. A, 2021).

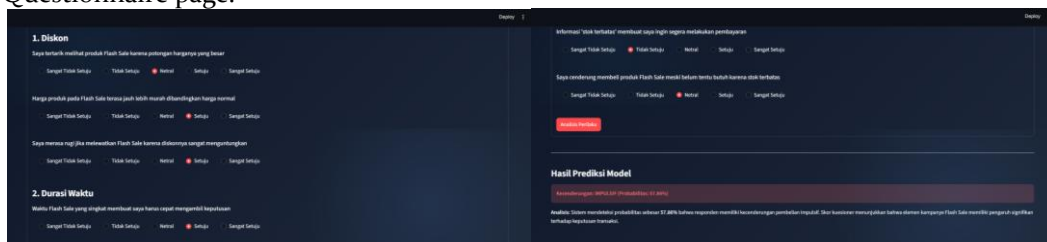
3.2 System Implementation

The developed web-based system successfully integrates the questionnaire completion process, dataset processing, model training, and visualization of classification results. The system consists of several main pages, namely:

1. System homepage

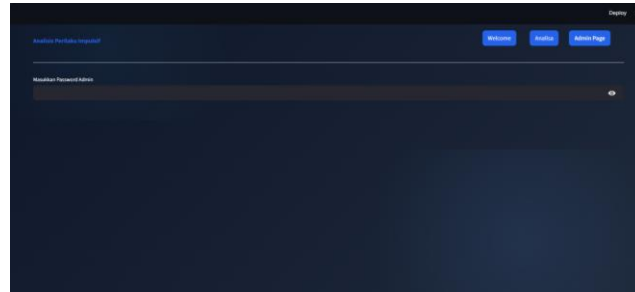


2. Questionnaire page.



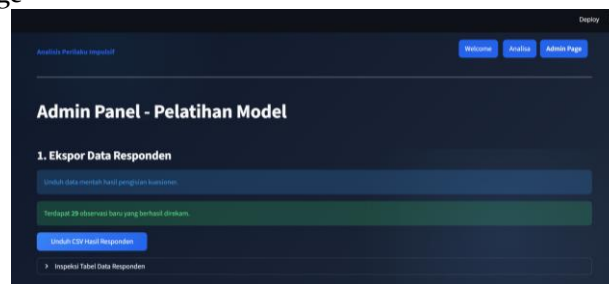
The questionnaire page is a key component of the system that serves as a medium for collecting respondent data.

3. Admin dashboard



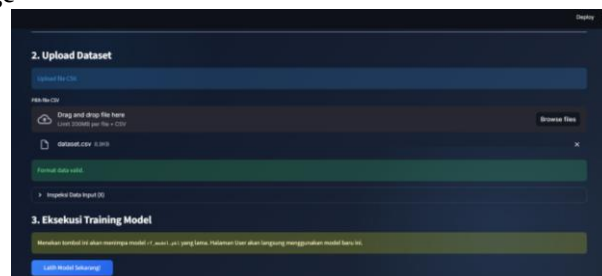
The admin page is used to manage the system, particularly for data processing and model training. To access this page, the admin must enter a password as a security measure.

4. Dataset upload page



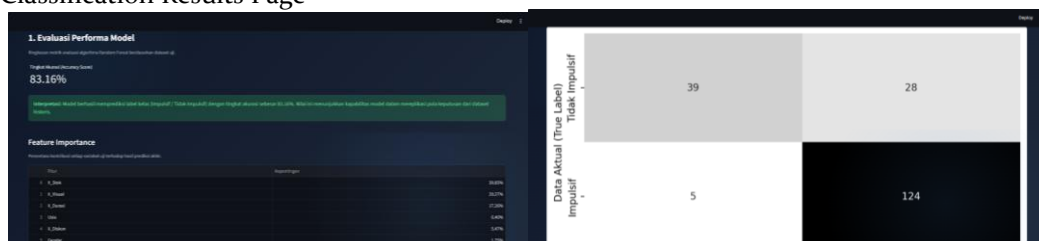
In the respondent data export section, all individuals who attempt to complete the Likert scale questionnaire in the first module will be automatically recorded in .csv format with a timestamp history and can also be downloaded.

5. Model training page



After the dataset has been successfully uploaded, the admin can initiate the model training process by clicking the “Train Model Now” button. This process will generate a Random Forest model that will be used by the system to make predictions.

6. Classification Results Page





Based on the results obtained, it appears that the limited stock variable (39.85%) has the greatest influence. A demographic distribution analysis was conducted to examine purchasing behavior trends by gender. Among men, 64.3% exhibited impulsive behavior and 35.7% exhibited rational behavior. Meanwhile, among women, 67.3% exhibited impulsive behavior and 32.7% exhibited rational behavior. These results indicate that both men and women tend to make impulsive purchases during flash sales, with the percentage of impulsive behavior among women being slightly higher than that of men.

System implementation shows that all features function properly and can be used to assist in the classification of impulsive behavior. The system was developed to automate the analysis of impulsive purchasing behavior. Through this system, users can process data and view classification results more quickly and efficiently. The website system is designed with a simple interface to make it easy to use for both users and administrators. Additionally, the system can assist in visualizing classification results, making the generated information easier to understand. The system was developed using the Python programming language and integrated with the Random Forest algorithm to support the automatic classification of impulsive purchasing behavior.

3.3 Classification Results

The Random Forest algorithm successfully classified respondents' impulsive purchasing behavior tendencies into impulsive and non-impulsive categories. Model evaluation results show that the Random Forest algorithm achieved an accuracy rate of 83.16%. This accuracy score indicates that the model performs well in identifying patterns of impulsive purchasing behavior based on flash sale features. The confusion matrix shows that the majority of the data was correctly classified into their respective categories. This proves that Random Forest is effective for use in machine learning-based consumer behavior research.

3.4 Feature Importance Analysis

One of the key advantages of Random Forest is its ability to identify the degree of influence each variable has on the classification results.

1. Discounts

Discounts were the factor that most influenced respondents' impulsive behavior. Significant price reductions made respondents feel they were getting a better deal, prompting them to make a purchase immediately. Many respondents also felt they were missing out if they missed a flash sale with high discounts.

2. Time Duration

The short duration of promotions creates psychological pressure on consumers. Countdown timers prompt respondents to make purchasing decisions quickly without rationally considering their needs. Time pressure has been shown to increase impulsive behavior among Gen Z consumers.

3. Visuals and Notifications

Promotional notifications and product visuals also influence consumer attention. Respondents tend to immediately open the Shopee app when they receive a flash sale notification. Eye-catching visuals also increase respondents' interest in the product.

4. Stock Availability

Information about limited stock triggers a fear of running out of the product or the fear of missing out (FOMO). As a result, consumers are driven to buy the product even though they don't actually need it yet.

3.5 Discussion

The results of this study indicate that flash sale features have a significant influence on Gen Z's impulsive purchasing behavior. Factors such as large discounts, time pressure, promotional notifications, and limited stock can trigger emotional responses in consumers, thereby reducing rational consideration in decision-making. These findings align with the Stimulus-Organism-Response (S-O-R) theory proposed by Mehrabian and Russell (1974), in which flash sale features act as external stimuli that influence consumers' psychological states and result in impulsive purchasing.

This study also demonstrates that machine learning algorithms such as Random Forest can be used to objectively classify consumer behavior. Compared to conventional statistical methods, Random Forest offers more flexible and stable classification capabilities. An accuracy rate of 83.16% indicates that Random Forest is effective for analyzing consumer behavior based on digital data. The developed web-based system can also facilitate the analysis of impulsive behavior more efficiently.

The findings of this study are expected to serve as a basis for evaluation by e-commerce platforms in implementing more responsible digital marketing strategies

4. Conclusion

This study successfully applied the Random Forest algorithm to classify Gen Z's impulsive purchasing behavior in relation to Shopee's flash sale feature. The findings show that flash sale characteristics significantly influence impulsive buying behavior among consumers. The model achieved an accuracy of 83.16%, indicating that Random Forest is an effective method for behavioral classification in e-commerce contexts. Among the examined variables, discount magnitude and promotion duration are the most influential factors, followed by visual/notification elements and stock availability. These results provide important practical implications for e-commerce strategy development. Platforms such as Shopee can optimize flash sale design by prioritizing discount intensity and carefully regulating promotion duration to create stronger urgency effects and increase conversion rates. Additionally, the use of attractive visual cues, real-time notifications, and limited stock indicators can further enhance consumer engagement and stimulate impulsive purchasing behavior. These insights can help digital marketers design more effective promotional strategies based on data-driven behavioral patterns, particularly for targeting Gen Z consumers who are highly responsive to digital stimuli. This study also contributes by developing a web-based classification system that enables automatic prediction and visualization of impulsive buying behavior, supporting real-time decision-making in e-commerce analytics. However, the study has limitations. The dataset size is relatively limited, which may reduce the generalizability of the model. In addition, some behavioral variables may be influenced by self-reported responses, which can introduce bias in representing actual purchasing behavior. The study also focuses only on the Random Forest algorithm without extensive comparison to other advanced machine learning models. Future research should expand dataset diversity, incorporate additional behavioral features, and compare multiple classification algorithms to improve model robustness and predictive performance in analyzing impulsive buying behavior.

References

- Aggarwal, P., Jun, S. Y., & Huh, J. H. (2011). Scarcity messages and consumer behavior. *Journal of Advertising*, 40(3), 19–30. <https://doi.org/10.2753/JOA0091-3367400302>
- Beatty, S. E., & Ferrell, M. E. (1998). Impulse buying: Modeling its precursors. *Journal of Retailing*, 74(2), 169–191. [https://doi.org/10.1016/S0022-4359\(99\)80092-X](https://doi.org/10.1016/S0022-4359(99)80092-X)
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Chan, T. K. H., Cheung, C. M. K., & Lee, Z. W. Y. (2017). The state of online impulse-buying research.

- International Journal of Information Management, 36(3), 204–216. <https://doi.org/10.1016/j.ijinfomgt.2015.11.001>
- Djafarova, E., & Bowes, T. (2021). Instagram and impulse buying in Gen Z. *Journal of Retailing and Consumer Services*, 59, 102345. <https://doi.org/10.1016/j.jretconser.2020.102345>
- Hajli, N. (2015). Social commerce constructs and consumer behavior. *Technological Forecasting and Social Change*, 94, 311–324. <https://doi.org/10.1016/j.techfore.2015.01.005>
- Inman, J. J., Winer, R. S., & Ferraro, R. (2014). The interplay among consumer decision factors. *Journal of Marketing*, 78(3), 109–121. <https://doi.org/10.1509/jm.13.0185>
- Kotler, P., Kartajaya, H., & Setiawan, I. (2021). *Marketing 5.0*. Wiley.
- Laudon, K. C., & Traver, C. G. (2021). *E-commerce: Business, technology, society*. Pearson.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Mai, N. T. T., Jung, K., Lantz, G., & Loeb, S. G. (2003). Impulse buying behavior study. *Journal of Retailing and Consumer Services*, 10(2), 83–91. [https://doi.org/10.1016/S0969-6989\(02\)00041-9](https://doi.org/10.1016/S0969-6989(02)00041-9)
- Priporas, C. V., Stylos, N., & Fotiadis, A. K. (2017). Generation Z consumer behavior. *Journal of Tourism Futures*, 3(3), 199–208. <https://doi.org/10.1108/JTF-12-2016-0045>
- Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters in random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3), e1301. <https://doi.org/10.1002/widm.1301>
- Rook, D. W. (1987). The buying impulse. *Journal of Consumer Research*, 14(2), 189–199. <https://doi.org/10.1086/209105>
- Verplanken, B., & Herabadi, A. (2001). Individual differences in impulse buying. *Journal of Economic Psychology*, 22(1), 71–99. [https://doi.org/10.1016/S0167-4870\(00\)00033-X](https://doi.org/10.1016/S0167-4870(00)00033-X)
- Zhang, X., Prybutok, V. R., & Koh, C. E. (2018). The role of urgency in online impulse buying. *Decision Support Systems*, 105, 1–11. <https://doi.org/10.1016/j.dss.2017.11.001>