PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 22, No. 1 March 2026 | DOI: 10.33480/pilar.v22i1.7893

APPLICATION OF THE RANDOM FOREST ALGORITHM IN PREDICTING IMPULSE BUYING BY CONSUMERS ON INDONESIAN MARKETPLACE APPLICATIONS

Ni Nyoman Cory Villdina*; Monica Julia Efata; Etania Frederika; Muhammad Sony Maulana

Sistem Informasi Kampus Kota Pontianak
Universitas Bina Sarana Informatika, Pontianak, Indonesia
www.bsi.ac.id
amorakireiangel0702@gmail.com*, efataefata708@gmail.com, etaniafrederika@gmail.com, sony@bsi.ac.id
(*) Corresponding Author

Abstract—*The development of e-commerce in Indonesia has driven an increase in impulsive buying behavior influenced by emotional and situational factors. This study aims to determine the important factors that influence impulsive buying by applying the Random Forest machine learning algorithm. Data were obtained from 628 online shopping application users aged 18-40 years in Indonesia. The analysis results indicated that the majority of respondents fall into the medium impulsive buying category (67.9%), with a model accuracy of 86.5%. The Hedonic Browsing (HB) variable had the greatest influence on impulsive buying behavior, followed by Utilitarian Browsing (UB) and Time Perspective (TP). These findings confirm that emotional aspects and situational conditions play a more dominant role than rational considerations in driving spontaneous shopping decisions.*

Keywords: *consumer behavior, e-commerce, impulsive buying, machine learning, random forest.*

Abstrak—*Perkembangan e-commerce di Indonesia mendorong meningkatnya perilaku belanja impulsif yang dipengaruhi oleh faktor emosional dan situasional. Penelitian ini bertujuan untuk menentukan faktor-faktor penting yang berpengaruh terhadap pembelian secara impulsif dengan menerapkan algoritma machine learning random forest. Data diperoleh dari 628 pengguna aplikasi belanja online berusia 18-40 tahun di Indonesia. Hasil analisis menunjukkan mayoritas responden berada pada kategori pembelian impulsif sedang (67,9%), dengan akurasi model sebesar* 86.5%,. *Variabel Hedonic Browsing (HB) berpengaruh paling dominan terhadap perilaku pembelian impulsif, diikuti oleh Utilitarian Browsing (UB) dan Time Perspective (TP). Temuan ini menegaskan bahwa aspek emosional dan kondisi situasional memiliki peran lebih dominan dibandingkan pertimbangan rasional dalam mendorong keputusan belanja spontan.*

Kata Kunci: *perilaku konsumen, pembelian impulsif, e-commerce, machine learning, random forest.*

INTRODUCTION

The development of the digital world has now penetrated almost all aspects of life, including people's shopping habits. This digital transformation has given rise to a new system where buying and selling activities can be carried out without any space or time limitations. The habit of shopping directly at physical stores has begun to shift towards online purchases through various e-commerce and marketplace platforms. Today, people are increasingly accustomed to using online shopping services to meet various needs, ranging from kitchen ingredients, clothing, cosmetics, and electronics, to shuttle services and goods delivery. These changes not only provide convenience for consumers, but also alter the nature of interaction between sellers and buyers in the digital space (Safaroh, 2023). This massive development is marked by the soaring value of Indonesian e-commerce transactions, placing it as one of the largest digital markets in Southeast Asia. This phenomenon has created an environment that is highly conducive to the emergence of shopping behavior driven by emotions and situations, which is the main focus of this study.

However, behind these conveniences, the phenomenon of impulsive buying emerges, which is the behavior of purchasing goods spontaneously without prior planning. According to (Rinonce, Jannah, Amelia, Anggun, & Prasetyo, 2025), this behavior can be triggered by internal factors such as momentary emotions, curiosity, temporary feelings of pleasure, or fear of missing out (FOMO). In addition, external factors such as flash sales, discounts, and product recommendations also have an influence. Research conducted by (Vidyanata, Junianto, & Setiobudi, 2024) shows that browsing e-commerce sites with hedonic motivation can increase the tendency for impulsive buying, especially among Generation Z. This shows that online purchasing decisions are often influenced more by emotional impulses than rational considerations. This impulsive shopping behavior is specifically influenced by situational and emotional variables, namely: Hedonic Browsing (HB), which emphasizes the pleasure of browsing; Utilitarian Browsing (UB), which focuses more on functional needs; and Time Perspective (TP), which is related to time, such as sudden offers and discounts. These variables do not operate independently, but interact dynamically in shaping impulsive buying behavior. The mechanism of impulsive buying is conceptualized as a dynamic interplay between functional objectives and emotional triggers. It typically originates from Utilitarian Browsing (UB), where consumers initially access the platform with goal-oriented tasks. However, engaging interface stimuli frequently transitions consumers into Hedonic Browsing (HB), a pleasure-seeking activity that heightens emotional satisfaction while diminishing rational constraints. This transition is further catalyzed by Time Perspective (TP) through features such as flash sales or limited-time offers, which instill a sense of psychological urgency. This temporal pressure compels consumers to bypass deliberate evaluation to avoid the 'fear of missing out' (FOMO), effectively transforming a rational shopping intent into spontaneous impulsive buying behavior. Understanding how these variables interact is crucial for predicting spontaneous shopping tendencies in Indonesian marketplaces.

Various previous studies have highlighted similar topics. Most studies use a descriptive statistical analysis approach, such as the study conducted by (Gozali & Pamungkas, 2025), which examined the influence of digital marketing on impulsive behavior among Generation Z TikTok users in Indonesia through multiple linear regression analysis. Although the multiple linear regression method is effective in measuring simple relationships, it faces significant challenges when used to analyze complex and non-linear consumer behavior data, and is less effective in predicting categorical results (such as low, medium, or high levels). Meanwhile, simple Decision Tree models (such as C5.0) tend to face overfitting problems and have difficulty producing consistent outputs when dealing with varied and complex datasets. Another study by (Wan, Zeng, & Zhang, 2024) analyzed online shopping addiction among college students using the C5.0 Decision Tree machine learning method. However, despite these previous studies, there is still a research gap in the use of more robust machine learning approaches to predict impulsive buying behavior, particularly in classifying the level of impulsive buying. Conventional statistical methods are generally designed to analyze relationships between variables rather than to perform predictive classification tasks, making them less effective in capturing complex and non-linear patterns in consumer behavior data. Therefore, more 7advanced machine learning approaches are required to improve prediction performance and provide more reliable classification results.

For the data mining process in this study, a machine learning approach using the Random Forest algorithm is applied. Random Forest is chosen due to its strong predictive performance, its ability to handle complex data relationships, and its capability to identify important variables through feature importance analysis (Ketipov, Angelova, & Doukovska, 2023). This algorithm also helps reduce the overfitting problem commonly found in single decision tree models. The use of Random Forest to predict the level of impulsive buying influenced by situational and hedonistic factors (HB, UB, TP) in Indonesian e-commerce is a major innovation in this study. Therefore, the objectives of this study are (1) to classify the level of consumer impulsive buying and (2) to assess the importance of situational and emotional factors in influencing this behavior. Thus, the results of this study are expected to enrich the understanding of modern consumer behavior and serve as a reference for e-commerce businesses in designing more efficient and targeted marketing strategies.

MATERIALS AND METHODS

A quantitative approach combined with predictive data mining techniques was applied in this study to examine and categorize impulsive buying behavior among Indonesian marketplace users. To determine the factors associated with this behavior, the Random Forest algorithm was selected as the main analytical tool. The analytical process followed the CRISP-DM framework, which structures the data mining workflow into six sequential stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

Business Understanding

This study applies the data mining method using the Random Forest algorithm to analyze the

Vol. 22, No. 1 March 2026 | DOI: 10.33480/pilar.v22i1.7893

factors that influence impulsive purchasing behavior in research datasets. The data mining method was chosen for its effectiveness in extracting patterns, relationships, and important information from data sets in a structured manner, with the aim of grouping, predicting, and recognizing individual behavior based on specific characteristics (Andini et al., 2024).

According to research conducted by (Wang & Lin, 2025), Random Forest is highly effective for processing large and complex data, providing high accuracy for analyzing the impact and relationship between variables, identifying the most significant variables for prediction results, and excelling in reducing the possibility of overfitting through a combination of various Decision Trees.

Data Understanding

The data source used comes from research conducted by (Budiman & Wijaya, 2023), by distributing an online questionnaire to eligible participants, namely Indonesian consumers aged 18-40 who own a mobile device. This dataset was selected because it has characteristics that are in line with the research objectives, namely to identify and predict the factors that influence impulsive buying on marketplace platforms.

The decision to use this secondary dataset stems from its proven reliability and close fit with our target demographic of Indonesian online shoppers. Because it is a peer-reviewed publication, the data's validity is already confirmed. From an ethical standpoint, the dataset is freely accessible for scholarly purposes. Furthermore, the original study adhered to standard protocols by anonymizing respondent data and obtaining proper informed consent.

This dataset includes survey results from 628 respondents who use online marketplace apps in Indonesia. To clarify the role of each variable in the predictive model, the six main variables are summarized based on their function as independent (predictors) and dependent (target) variables, as shown in Table 1:

Table 1. Summary of Research Variables

Variable Role	Variable Name and Detailed Description
Independent Variables (Predictors)	Physical Environment (PE) The visual appeal and layout of the app, such as attractive user interfaces (UI), vibrant promotional banners, or interactive animations.
	Social Environment (SE) Shopping is influenced by social factors, such as peer recommendations, influencer endorsements, or seeing high ratings and positive reviews from other buyers.
PE (3)	Time Perspective (TP) Time-related constraints and triggers that create a sense of urgency, such as flash sale countdown timers or limited-time midnight discounts.
PE (4)	Utilitarian Browsing (UB) Goal-oriented browsing based on functional needs.
SE (1)	Hedonic Browsing (HB) Browsing driven by entertainment and pleasure without a specific purchase intention.
Dependent Variable (Target)	Impulsive Buying (IB) The actual behavior of making spontaneous, unplanned purchases is triggered by sudden desires or the situational factors mentioned above.

Source: (Research Results, 2025)

This dataset contains respondents' statements on a scale of 1-5, where 1 = “Strongly disagree,” 2 = “Disagree,” 3 = “Neutral,” 4 = “Agree,” and 5 = “Strongly agree.” To determine the level of Impulsive Buying in Indonesia (low/medium/high), the method is to interpret the average score range by subtracting 5 from 1, then dividing it into 3 parts to obtain an interval of approximately ≈1.33. Thus, the average score for the low category is (≤ 2.33), medium (> 2.33 - ≤ 3.67), and high (> 3.67), according to the interval division (Oktavia, 2022).

The data Understanding stage begins with the Dataset Overview and Data Preprocessing, which involves importing the dataset into Python to check and detect any missing values, duplicate data, or input errors that could affect the analysis results, as well as ensuring that all variables can be processed.

Data Preparation

In the data preparation stage, several preprocessing steps were carried out to ensure the quality and consistency of the dataset before the modeling process. These steps included data cleaning to check for incomplete or inconsistent data and data transformation to prepare the variables for further analysis.

After the preprocessing stage, the dataset was divided into training and testing sets using the train–test split technique. Based on previous research (Fahrudin, Putra, & Umaroh, 2024), several data distribution ratios can be used, such as 80:20, 70:30, and 60:40, depending on the size of the dataset and the research objectives.

Previous studies (Sivakumar & Parthasarathy, 2024) suggest that using a 70:30 ratio may reduce model accuracy due to limited training data, while a 90:10 ratio may increase the risk of bias because the testing data becomes too small. Therefore, this study uses an 80:20 ratio to provide a balanced distribution between training and testing data.

The training data is used to train the model to recognize the relationships between variables (PE, SE, TP, UB, and HB) and impulsive buying behavior (IB). Meanwhile, the testing data is used to evaluate the model's ability to make predictions on unseen data.

Modeling

The Modeling stage was conducted to build a prediction model using the Random Forest algorithm as a combination of several predictor trees (Decision Trees) to produce a more stable and accurate prediction for predicting impulsive buying based on variables, and aiming to identify the relationship between variables.

To ensure reproducibility and optimal performance, specific hyperparameters were configured during the training process. The model was built using 100 trees (n_estimators = 100), and the random_state parameter was set to 42 to guarantee consistent results across multiple executions. Meanwhile, the tree depth (max_depth) was kept at its default unrestricted value to allow the nodes to expand optimally.

Evaluation

Model evaluation is used to assess how well the model predicts impulsive purchasing behavior. The evaluation is conducted using a confusion matrix as the basis for measuring classification performance. Several metrics are reported to provide a comprehensive assessment, including Accuracy, Precision, Recall, and F1-score (Ramadhan, Abadi, M S, & Yudiarso, 2024).

Accuracy measures the overall correctness of predictions. Although it is reported for completeness, it may not fully reflect the model's ability to predict minority classes due to the class imbalance in the dataset (e.g., low impulsive buying is underrepresented). Therefore, additional metrics such as Precision, Recall, and F1-Score are employed to provide a more comprehensive evaluation. Precision measures the accuracy of positive predictions made by the model, as formulated in Equation (2). Recall evaluates the model's ability to identify all actual positive instances, as presented in Equation (3). Furthermore, the F1-Score represents the harmonic mean of Precision and Recall, aiming to balance both metrics, as shown in Equation (4). The metrics are calculated as follows (Obi, 2023):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Description:
TP (True Positive) = This metric represents instances where the model successfully identifies positive cases as positive.
TN (True Negative) = This occurs when the system accurately recognizes negative data points and classifies them correctly.
FP (False Positive) = A False Positive occurs when the classification model mistakenly identifies a negative data point as being positive.
FN (False Negative) = This value indicates cases where positive data is incorrectly dismissed or categorized as negative by the model.

Deployment

This phase marks the concluding step, concentrating on the interpretation of the modeling findings. Instead of acting as a direct application of a real-world system, this stage is dedicated to analyzing and interpreting the results generated by the developed model. The analysis utilized the Python programming language, along with several supporting libraries such as pandas and NumPy for data manipulation, and scikit-learn for building and evaluating the Random Forest model. In this phase, a Feature Importance evaluation is carried out to identify the factors that have the most significant impact on predicting impulsive purchasing behavior. The results of the analysis can be illustrated using bar or pie charts to represent each factor in relation to the predictions, as well as graphs that depict the overall effectiveness of the model, enhancing clarity and understanding of the influence of each factor (Yasin, Ding, Mamat, Guo, & Song, 2025).

RESULTS AND DISCUSSION

This section presents the results of data analysis on factors that influence impulsive buying among online market users. This study uses a questionnaire containing statements based on indicators for each variable. Each statement is evaluated using a 5-point likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

As this study employs a secondary dataset from Budiman & Wijaya (2023), the validity and reliability of the questionnaire instruments were already verified by the original researchers, ensuring that variables such as PE, SE, TP, UB, and HB are robust for further analysis. Regarding the data distribution, although the 'Low' category represents only 3.5% of the total samples, we decided to retain the original distribution to preserve ecological validity. In the context of Indonesian e-commerce, a balanced dataset would be an artificial construct that fails to reflect the reality of widespread impulsive consumption. By avoiding synthetic oversampling (SMOTE), we ensure that the model learns from authentic psychological responses rather than artificial patterns.

Random Forest was specifically selected due to its Bootstrap Aggregating (Bagging) mechanism. This ensemble approach builds multiple decision trees on varied subsamples, which naturally mitigates the risk of the model being biased toward the majority class. This allows the algorithm to maintain high predictive stability even when faced with skewed class distributions, ensuring that the feature importance results remain reliable and representative of actual consumer behavior.

Details of the questionnaire items for each variable can be seen in the following Table 2.

Table 2. Variables and statements in datasets

Variable	Items
PE (1)	I find that using this online marketplace makes my shopping tasks much more efficient and convenient.
PE (2)	The overall look and visual layout of the marketplace app I use is quite professional and well-crafted.
PE (3)	I am particularly satisfied with how visually appealing the interface of this shopping platform is.
PE (4)	Everything on this platform is laid out in a way that is very easy for me to understand and navigate.
SE (1)	Most people I know believe that adopting an online marketplace app is a smart move.
SE (2)	My social circle generally supports and encourages the idea of using these online shopping platforms.
SE (3)	Personal suggestions from my friends have been a major factor in why I use online marketplace apps.
TP (1)	The app is highly reliable because it works perfectly for me, no matter where I happen to be.
TP (2)	I can use this marketplace app without any technical issues, whether I am at home or in my workplace.
TP (3)	I feel completely confident using the platform because its features remain stable regardless of my location.
UB (1)	I looked through apps on the online marketplace to find items that cost less.
UB (2)	I looked through apps on the online marketplace to find better quality items to buy.
UB (3)	My time spent on various marketplace platforms is often dedicated to the systematic collection of product details and relevant information.
UB (4)	I browsed different apps on the online marketplace to compare the stores.
UB (5)	I looked for apps on the online marketplace board that can make online shopping more efficient.
HB (1)	When I feel like I can forget about my troubles while browsing online marketplace applications.
HB (2)	I feel like I can enjoy my downtime while browsing online marketplace applications.
HB (3)	I really enjoyed looking through the online marketplace board applications.
IB (1)	Navigating through marketplace platforms often triggers a spontaneous desire to acquire supplementary products that exceed my original purchasing intent.
IB (2)	During the browsing process, I frequently experience a strong impulse to purchase goods that are entirely disconnected from my primary shopping objectives.
IB (3)	I often end up buying things that aren't on my shopping list when I'm just browsing.

Source: (Budiman & Wijaya, 2023)

The questionnaire data from respondents was compiled into Microsoft Excel and imported into Google Collab. Next, data checking and cleaning were carried out to ensure that no values were missing, so that the data was accurate and all variables were ready for use.

After that, proceed to the Train-Test Split stage. There are a total of 502 data points (80%) were used as training data to build the model, and 126 data points (20%) were used as testing data to evaluate the performance of the random forest model, as shown in the image below.

```
# ====================================================
# STEP 6: Prepare feature data (X) and target data (y)
# ====================================================

# Use PE, SE, TP, UB, HB as independent variables (situational factors)
# Use Kategori_IB as the dependent variable (target)
X = df[['PE','SE','TP','UB','HB']]
y = df['Category_IB']

# Encode the target so it can be used in the model
le = LabelEncoder()
y_enc = le.fit_transform(y)

# Divide the data into training and testing
X_train, X_test, y_train, y_test = train_test_split(
    X, y_enc, test_size=0.2, random_state=42, stratify=y_enc)

print("Training data:", len(X_train))
print("Testing data:", len(X_test))

Training data: 502
Testing data: 126
```

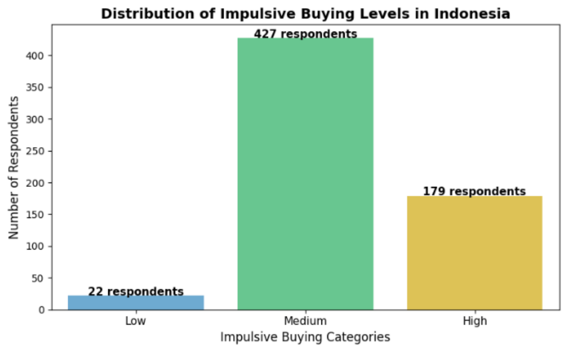
Source: (Research Results, 2025)
Figure 1. Train-Test Split Data Training and Testing

After dividing the data into training sets and testing sets, and also training the model using the Random Forest algorithm, the next step is to analyze the distribution of impulsive buying levels from the processed respondent data. After running the program code in Figure 2, it will display a visualization as shown in Figure 3.

```
# Hitung jumlah responden per kategori
kategori_counts = df['Kategori_IB'].value_counts().reindex(["Rendah", "Sedang", "Tinggi"])
kategori_counts = kategori_counts.fillna(0)  # jika ada kategori yang kosong

# --- DIAGRAM BATANG (BAR CHART) ---
plt.figure(figsize=(8,5))
sns.barplot(x=kategori_counts.index, y=kategori_counts.values, palette=["#5DADE2", "#58D68D", "#F4D03F"])

# Tambahkan label di atas tiap batang
for i, val in enumerate(kategori_counts.values):
    plt.text(i, val + 0.5, f"{int(val)} responden", ha='center', fontsize=11, fontweight='bold')

plt.title("Distribusi Tingkat Impulsive Buying di Indonesia", fontsize=14, fontweight='bold')
plt.xlabel("Kategori Impulsive Buying", fontsize=12)
plt.ylabel("Jumlah Responden", fontsize=12)
plt.xticks(fontsize=11)
plt.yticks(fontsize=10)
plt.tight_layout()
plt.show()
```

Source: (Research Results, 2025)

Figure 2. Code program for visualizing the distribution of impulse buying levels



Percentage of Impulsive Buying in Indonesia

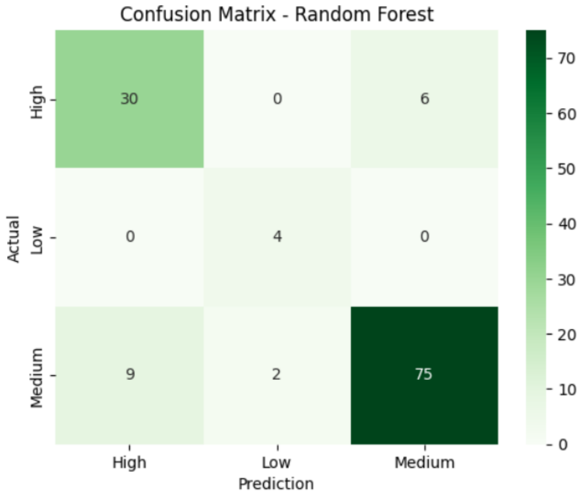
Source: (Research Results, 2025)

Figure 3. Distribution of Impulsive Purchasing levels in Indonesia

Based on Figure 3, the results of the analysis of the 628 respondents, it shows that most respondents are in the medium impulsive buying category, namely 427 respondents (67.9%), 179 respondents (28.5%) in the high category, and 22 respondents (3.5%) in the low category. This shows that impulsive shopping behavior among Indonesian consumers is at a medium level, meaning that most consumers make spontaneous purchases but still consider rational aspects before making a decision.

Following the identification of respondents' impulsive buying level distribution, the next step is to measure the effectiveness of the Random Forest model through a confusion matrix that illustrates the relationship between the predicted outcomes and the actual data. The confusion matrix classifies impulsive buying levels into three categories: low, medium, and high.

Based on the confusion matrix generated in Figure 4, the model demonstrates a robust performance in recognizing general patterns of impulsive buying, although some misclassifications persist due to the inherently unbalanced data distribution. In the low category, while the sample size is limited, the model correctly predicted all 4 actual respondents without misclassifying them into higher classes. This stability in the minority class suggests that the Random Forest's Bagging mechanism effectively preserves the characteristics of low-impulse behavior despite the potential for the model to be biased toward the majority distribution.This condition shows that the model tends to pull the class category to a higher class, due to the similarity of data patterns between the two categories.



Source: (Research Results, 2025)

Figure 4. *Confusion Matrix*

For the medium category, the model results were significantly more consistent, with a total of 75 respondents correctly predicted according to their original labels. However, minor errors occurred where 2 respondents were predicted as low and 9 as high. Despite these few misclassifications, the model's performance in this category is the most stable, driven by the larger volume of data which allowed for a more optimal learning process and better generalization of moderate impulsive buying patterns.

In the high category, the model also performed well, successfully identifying 30 respondents. However, 6 respondents were incorrectly predicted as medium category. The dominant error occurred due to this shift to the medium class, signifying 'fuzzy boundaries' in consumer psychology. This indicates an overlap in characteristics between high and medium impulsivity, where emotional triggers often share similar behavioral traits in a digital marketplace context.

Overall, the most accurate predictions are found in the medium category, followed by the high category with strong correct classifications. The weakest performance appears in identifying the low category due to its small representation, yet the model still managed to capture all actual instances. In general, this model provides reliable

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 22, No. 1 March 2026 | DOI: 10.33480/pilar.v22i1.7893

predictions by maintaining ecological validity, reflecting the actual skewed impulsive buying patterns in Indonesia without the need for artificial data adjustments.

After obtaining the confusion matrix evaluation results, the next step is to analyze the precision, recall, and F1-score values to evaluate the model's performance. The following is the classification report results from the Random Forest model shown in the image below.

```
=== Random Forest ===
Accuracy: 0.8650793650793651
              precision    recall  f1-score   support

        High       0.77      0.83      0.80        36
         Low       0.67      1.00      0.80         4
      Medium       0.93      0.87      0.90        86

    accuracy                           0.87       126
   macro avg       0.79      0.90      0.83       126
weighted avg       0.87      0.87      0.87       126
```

Source: (Research Results, 2025)

Figure 5. *Classification Report*

Based on the Figure 5, for the low category, the model shows relatively low performance with a precision of 0.67, a recall of 1.00, and an F1-score of 0.80. These values indicate that although the low category contains only a few respondents, the model is still able to correctly identify all actual low-category instances. The high recall suggests that the model’s sensitivity is well-preserved, even if the small sample size naturally limits the precision of the predictions.

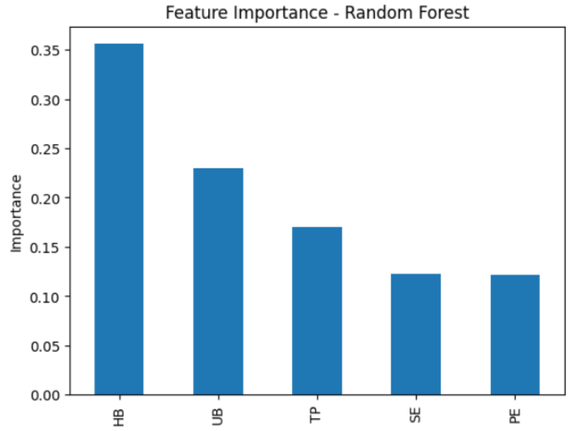
In the medium category, the results were significantly more stable, with a precision value of 0.93, a recall value of 0.87, and an F1-score of 0.90. These values were the highest among the three classes, indicating that the model was best at recognizing patterns of respondents with moderate levels of impulsive buying. This high F1-score confirms that the Random Forest algorithm is highly effective at generalizing behavior when provided with sufficient training data, capturing the core characteristics of the most prevalent consumer group in the dataset.

Meanwhile, in the high category, the model achieved a precision of 0.77, a recall of 0.83, and an F1-score of 0.80. The high recall value indicates that most respondents with a high tendency toward impulsive buying were successfully identified by the model. The slight dip in precision here further reflects the 'fuzzy boundaries' between moderate and high impulsivity, as consumers in these groups often respond to similar psychological triggers in an e-commerce environment.

Overall, an accuracy value of 0.87 indicates that the model is capable of providing correct predictions for approximately 87% of all test data. The macro average F1-score of 0.83 shows a slight imbalance in performance between classes, with the low category being statistically more challenging to predict. However, the weighted average F1-score of 0.87 indicates that the model provides stable performance when considering the actual proportion of data in each class. This confirms that the Random Forest model works quite well for the Indonesian e-commerce context, proving its ability to generalize complex consumer behaviors even when faced with skewed data distributions.

After obtaining an overview of the classification model's performance evaluation, the next step is to understand the internal factors that influence the model's predictions using feature importance. The higher the importance value, the greater its role in influencing the classification results. Visualization can help identify which variables are dominant.

```
Feature Importance:
HB    0.355907
UB    0.229903
TP    0.170234
SE    0.122324
PE    0.121631
dtype: float64
```



Source: (Research Results, 2025)

Figure 6. Feature Importance

From the results of the feature importance analysis in Figure 6, it can be seen that the Hedonic Browsing (HB) variable has the greatest contribution in influencing impulsive buying behavior, followed by Utilitarian Browsing (UB), Time Perspective (TP), Social Environment (SE), and Physical Environment (PE). This shows that emotional and situational aspects play an important role in encouraging spontaneous or impulsive buying.

The dominance of the Hedonic Browsing (HB) factor indicates that there has been a transformation of the marketplace platform. The application no longer functions merely as a store for finding necessary items, but has evolved into a means of seeking entertainment. This suggests that the emotional satisfaction gained from the browsing process itself is now the main driver for spontaneous purchases. Features like continuous scrolling or personalized product recommendations are specifically designed to keep users engaged, making the pleasant experience of 'looking around' more powerful in triggering buying than the actual need for a product.

This explains why Hedonic Browsing (HB) has a greater influence than functional factors such as Utilitarian Browsing (UB). Meanwhile, the lower influence of Physical Environment (PE) and Social Environment (SE) indicates that in the current digital context, visual aesthetics and social pressure are no longer the primary triggers. Since most e-commerce platforms now offer similar high-quality interfaces, the visual environment is seen as a standard and no longer surprises the user. Furthermore, the low impact of social influence suggests that modern consumers are more independent in their browsing habits, making decisions based on their own personal impulses rather than being influenced by peer recommendations or external social cues.

Based on the entire performance evaluation analysis and classification matrix, the model used shows that the model's performance is effective and adequate in identifying and predicting impulsive buying behavior, with Hedonic Browsing as the main factor.

CONCLUSION

The results of this study indicate that impulsive buying behavior among Indonesian consumers generally falls within the medium category, with approximately 67.9% of the 628 respondents showing a tendency to make spontaneous purchases while still considering rational aspects. The Random Forest model developed in this research achieved an accuracy of 86.5%, indicating strong performance in classifying the level of impulsive buying behavior. By maintaining the original skewed data distribution without artificial resampling, these findings demonstrate that consumer impulsivity in Indonesian e-commerce environments can be effectively and authentically analyzed using machine learning approaches.

Furthermore, the analysis reveals that Hedonic Browsing (HB) is the most influential factor affecting impulsive buying behavior, followed by Utilitarian Browsing (UB) and Time Perspective (TP). These findings indicate that emotional and situational factors play a more dominant role in shaping impulsive buying decisions compared to purely rational considerations. The dominance of HB signifies that e-commerce platforms have evolved into hedonic entertainment hubs, where the browsing experience itself acts as the primary trigger for spontaneity. Social Environment (SE) and Physical Environment (PE) were also found to contribute to impulsive behavior, although their influence is relatively smaller. This reflects a standardization effect in modern e-commerce, where digital aesthetics and social cues have become standard expectations rather than unique competitive drivers.

This study contributes to the advancement of consumer behavior research by demonstrating the applicability of machine learning techniques, particularly the Random Forest algorithm, in predicting impulsive buying behavior in Indonesian e-commerce contexts. The findings also provide practical implications for e-commerce platforms and digital marketers, as insights into browsing behavior and time-related stimuli can support the development of more targeted promotional strategies that prioritize enjoyable and personalized browsing experiences.

However, this study has several limitations. The distribution of impulsive buying categories may contain potential class imbalance, and the dataset used in this research may not fully represent all segments of Indonesian online consumers. In addition, the study focuses on a single machine learning approach. Further research is suggested to assess the capability of the Random Forest algorithm in comparison with other machine learning approaches and to incorporate larger and more heterogeneous datasets so that predictive models of impulsive buying behavior can become more robust and broadly applicable in the dynamic digital economy.

REFERENCE

Andini, E. A., Maharani, B. A., Apriliani, M. D., Tsabita, K., Pratyasanto, A. A., Keuangan, P., & Stan, N. (2024). ANALISIS TINGKAT PERILAKU KONSUMTIF BELANJA ONLINE DENGAN ALGORITMA K-MEANS (STUDI KASUS MAHASISWA XXX). *Integrative Perspectives of Social and Science Journal (IPSSJ)*, *2*(1).

Budiman, S., & Wijaya, T. (2023). Mobile app impulsive buying: A situational factors dataset analysis. *Data in Brief*, *50*, 109559. https://doi.org/10.1016/j.dib.2023.109559

Fahrudin, N. U. R. F., Putra, K. R., & Umaroh, S. (2024). *Influence of Data Scaling and Train / Test Split Ratios on LightGBM Efficacy for Obesity Rate Prediction*. *9*(2), 220–234.

Gozali, I., & Pamungkas, H. S. (2025). Impulsive Buying Among Tiktok Users (Study on Gen Z). *Digital Innovation: International Journal of Management*, *2*(3), 210–218. Retrieved from https://doi.org/10.61132/digitalinnovation.v2i3.464

Ketipov, R., Angelova, V., & Doukovska, L. (2023). *Predicting User Behavior in e-Commerce Using Machine Learning Predicting User Machine Learning Behavior in Using*. (September). https://doi.org/10.2478/cait-2023-0026

Obi, J. C. (2023). *A Comparative Study of Several Classification Metrics and Their Performances on Data A Comparative Study of Several Classification Metrics and Their Performances on Data*. (February). https://doi.org/10.30574/wjaets.2023.8.1.0054

Oktavia, S. N. (2022). EFL students' readiness to have online learning during the COVID-19 pandemic (A case at the even semester students of English Education Study Program of Sultan Agung Islamic University in the academic year 2021/2022). *Undergraduate Thesis*.

Ramadhan, D. R., Abadi, F. M., M S, F. M., & Yudiarso, A. (2024). *Multi Metode Dan Multi Perspektif Konstruk Budaya Kesenangan (Indulgence) Terhadap Pembelian Impulsive Online Dengan Machine Learning*. 145–164.

Rinonce, E. M., Jannah, M., Amelia, R., Anggun, Z., & Prasetyo, R. (2025). Fear of Missing Out Fuels Impulsive Buying Behavior in Gen Z. *Psikologia : Jurnal Psikologi*, *10*(1), 97–110. https://doi.org/10.21070/psikologia.v10i1.1847

Safaroh, N. (2023). Pengaruh Fear of Missing Out (FoMO), Shopping Enjoymen, dan Hedonic Shopping Motivation Terhadap Impulse Buying di E-Commerce Shopee Pada Waktu Flash Sale. *Journal of Economics, Business, &Entrepreneurship*, *5*, 34–38. https://doi.org/10.29303/alexandria.v5iSpecialIssue.604

Sivakumar, M., & Parthasarathy, S. (2024). *Trade-off between training and testing ratio in machine learning for medical image processing*. 1–17. https://doi.org/10.7717/peerj-cs.2245

Vidyanata, D., Junianto, Y., & Setiobudi, A. (2024). Hedonic browsing behaviour and its impact on impulsive buying among generation z. *Primanomics: Jurnal Ekonomi & Bisnis*, *22*(2), 34–43.

Wan, X., Zeng, J., & Zhang, L. (2024). Predicting online shopping addiction: a decision tree model analysis. *Frontiers in Psychology*, *15*(January). https://doi.org/10.3389/fpsyg.2024.1462376

Wang, T., & Lin, J. (2025). *Machine Learning – Based Random Forest Prediction of Online Shopping Behavior in the Digital Economy*. 315–320. https://doi.org/10.1145/3745133.3745186

Yasin, P., Ding, L., Mamat, M., Guo, W., & Song, X. (2025). *Machine Learning-Based Interpretable Screening for Osteoporosis in Tuberculosis Spondylitis Patients Using Blood Test Data : Development and External Validation of a Novel Web-Based Risk Calculator with Explainable Artificial Intelligence (XAI)*. (May), 2797–2821.