

## **DETERMINANTS AND CLASSIFICATIONS OF ONLINE SHOPPING CONSUMERS' PURCHASE INTENTION IN INDONESIA**

**Reninta Dewi Nugraheni\***

*Research Centre for Economics of Industry, Services, and Trade, National Research and  
Innovation Agency, Indonesia*

### **ABSTRACT**

Online shopping is still growing significantly, which is an upward trend for e-commerce platforms. It has grown in popularity among consumers for a number of reasons, driving Indonesian consumer behavior. This study aims to identify determinants and classifications of Indonesian online shopping consumers' purchase intention using the decision tree methodology. The notion underpinning this study is the phenomena of online shopping as a consumer behavior transition in Indonesia which has increased gradually over the last decades. The data was collected using a structured questionnaire and 820 respondents participated in this study. Then, the data was analyzed using C5.0 and CART 5.0 algorithms. The results show that the most influential determinants of the frequency of online shopping include the consumers' occupation, choice of e-commerce platforms, product origin, and product quality. The influence of several factors on customers' intentions to engage in shopping online can be determined, including job status, e-commerce platform, product sources, and product quality.

**JEL:** L81, L86, Q5.

**Keywords:** *e-commerce, consumer behavior, online shopping.*

### **1. INTRODUCTION**

In this age of technology and information advances, internet adoption in Indonesia has skyrocketed. Internet penetration was 64.8% in 2018, and it was predicted to increase to 73.7% in 2019-2020. There was a total of 212.35 million people used the internet, with a penetration rate of 76.8%, and it would reach 77.02% of the total population in 2022. This figure demonstrates that domestic internet users have increased and become a habit in daily life. The Indonesian Internet Providers Association (*Asosiasi Penyelenggara Jasa Internet Indonesia* (APJII)) estimated that most of Indonesians used the internet for social media (98.02%), public services (84.90%), and online transactions (79%) (APJII, 2022).

According to the Bank Indonesia Annual Meeting 2021 report, Indonesia recorded the total amount of online shopping transactions in e-commerce platforms of IDR 403 trillion in 2021 – growing 51.6% from IDR 266 trillion in the previous year. Meanwhile, the total online shopping transactions as of the end of November 2022 reached IDR 435 trillion, with a target value of IDR 489 trillion for the online shopping transactions in 2022. This indicated that it had achieved 88.95% of the target and this figure is expected to skyrocket in the future years. This condition suggests that the use of internet for online shopping is increasing significantly, which is a positive signal for the e-commerce platforms.

The rapid growth of internet in Indonesia has shifted consumer purchasing behaviors, from traditional to internet-based. Over the last decade, the online shopping has grown in popularity (Lim, Osman, Salahuddin, Romle, & Abdullah, 2016; Vasic, Kilibarda, & Kaurin, 2019; Vijay,

---

\* Corresponding Author: [reninta.dewi.nugraheni@brin.go.id](mailto:reninta.dewi.nugraheni@brin.go.id)

Received: 17-01-2023, Accepted: 27-12-2023, Published: 03-04-2024

P-ISSN: 2087-9954, E-ISSN: 2550-0066. DOI: <http://dx.doi.org/10.26418/jebik.v13i1.61399>

Prashar, & Sahay, 2019). Most of the consumers prefer the online shopping for its conveniences (Vyas & Bissa, 2017), such as time savings (Giningroem, Setyawati, & Wijayanti, 2022), and having a wide range of products (Meeprom & Silanoi, 2020) and lower prices (Lee & Chen-Yu, 2018). The online shopping begin with searching the products, comparing the prices or features, selecting the desired products, placing orders, and making payments (Sinha & Kim, 2012).

Although the consumers may not immediately shift from the offline to online shopping, they do use the internet as part of their purchasing process. Rehman, Bashir, Mahmood, Karim, & Nawaz (2022) & Saleem, Aslam, Kim, Nauman, & Khan (2022) discovered groups of online shopping consumers at varying stages and levels of adoption. Several consumers have one of three preferences as they deal with a purchase: (1) using the internet solely to browse product information and make in-store purchases; (2) browsing product and price information on the internet and completing the purchase online; or (3) never using the internet in the shopping process. Meanwhile, Rehman & Ansari (2023) also added that Group (1) and Group (2) tend to use online customer reviews to assess and evaluate online stores.

The consumer behavior influences the purchase decision and has an impact on marketing strategies, purchasing processes, products, and services offered over the internet. The consumers usually show an interest in a product sold online before deciding to purchase it. Their purchase intention has a significant impact on the purchase decision (Pasharibu, Soerijanto, & Jie, 2020). However, because the online shop consumers cannot directly see the products, their intention is not the only driver of purchase decision. The purchase intention refers to a consumer's willingness to purchase and is broadly defined as the consumer's ability to purchase the products or services (Ali & Sohail, 2018). According to Beck & Kenning (2015), the consumer behavior in purchasing online products can be predicted by their intentions. Gopinath (2019) defined the purchase intention as a consumer's plan to decide where to buy the products. In the context of online shopping, online shopping intention is the willingness of consumers to buy specific products over the internet (Ha, Nguyen, Pham, & Nguyen, 2021). During the browsing stage, the consumers can predict their chance of purchasing a product, which can lead to a customized online purchasing process in which they merely examine the product information to reduce the risk of shopping online. Alsoud & Othman (2018) has revealed that the online shopping intention is influenced by website credibility, website quality, security protection, after-sale service on the online shopping.

Several studies have evaluated the consumers' online shopping behavior from a variety of perspectives. Beck & Kenning (2015) found that sellers' reputation and the level of consumer trust influenced the consumers' online shopping intention. Ha *et al.* (2021) identified several factors influencing individual shopping intentions, including attitude, subjective norm, perceptions of behavioral control, perceptions of usefulness, and trust. In contrast, they found that perceived risk of online shopping had a negative impact on the online shopping intention. The perceived risk has been proven to have the greatest influence on the online purchasing intention among other factors. The consumers frequently used the internet for online shopping because it benefits them (Ha *et al.*, 2021). Further, Wahyudin, Yuliando, & Savitri (2020) revealed that consumer comfort and convenience in making transactions, particularly in the payment and delivery services, had a positive and significant effect on the consumers' attitude and behavioral intention in the online shopping.

Determinants of the purchase intention are one research topic that can be thoroughly investigated. An intriguing research topic during the past decade has been the analysis of the

consumer behaviour and intention in the online shopping. As a result, considering that the consumer behavior is the primary predictor, it is vital to investigate the influence of determinants of the online shopping attitude and intention. For this reason, more empirical researches on the consumers' behaviour and attitude in dynamic digital settings are required. Therefore, this study aims to identify the determinants and classifications of Indonesian online shopping consumers' purchase intention using the decision tree methodology with the C5.0 and Classification and Regression Tree (CART) algorithms. Results of this study are expected to contribute to the literature in predicting the online shopping consumers' behaviour.

## **2. THEORETICAL FRAMEWORK AND EMPIRICAL STUDIES**

Nowadays, the consumers have the option of shopping indirectly (online shopping or through mail order catalogs) or directly from stores. Theory of planned behavior (TPB) explains that it is not only individual consumers' attitudes that influence their online shopping intention, but also the attitudes of others around them and the availability of technology (Aliedan, Elshaer, Alyahya, & Sobaih, 2022). There are consumers who prefer the online shopping and some who oppose it. In short, the consumer attitudes toward the online shopping, others' opinions, and the availability of technology all play a role in shaping the online shopping intention. It is possible to accurately forecast diverse behavioral intentions by considering the behavioral attitude, subjective norm, and perceived behavioral control. These intentions, along with the perceived behavioral control, were found to play a significant role in explaining the variability observed in actual behaviors (Gu & Wu, 2019; Sussman & Gifford, 2019).

Furthermore, the purchase intention is one of the determinants of consumers' shopping behavior, and it is strongly influenced by the individual and perceived behavioral control. The consumers' purchase intention is influenced by their search for pre-purchase information, which can be found in the online consumer reviews. These online reviews can be either positive or negative (Prasad, Garg, & Prasad, 2019). The recognition of products sold online can be achieved gradually and continuously when the consumers have developed trust during the online shopping process. Online transactions can be thought of as a process of gathering and exchanging information, and purchasing products. Individuals who own several accounts in web-based social media platforms develop a strong sense of self-concept and react rapidly. Their reactions also often influence their purchase intention. Previous studies indicated that if people are sensitive to or influenced by comments and posts on the social media positively, they will likely develop brand loyalty (Chua & Chang, 2016; Kardinasari, Iskandar, Nugraha, & Jatnika, 2019).

Purchasing a product in the online shopping refers to the act of selecting one or more available options. The consumers' attitude shaping the subjective norm has a direct impact on the purchase decisions (Truong, 2018). Several product evaluation processes are involved in the purchase decision, including analyzing the information received from any platform, fellow consumers, and real-world evaluation of product type, brand, price, quality, payment time, and payment method (Lu, Chang, & Chang, 2014). Indicators, such as product type, model, price affordability, product quality, purchase frequency, and payment method, have been used to measure the purchasing decision. This practice allows the consumers to obtain better product knowledge, which leads to increased satisfaction of online shopping (Karimi, Holland, & Papamichail, 2018).

Several algorithms, including C4.5, C5.0 and CART 5.0, have been widely discussed in the literature. Šebalj, Franjković, & Hodak (2017) developed a model predicting the shopping intentions. The model classified the respondents into two categories: those who planned to shop and those who did not. All models, with the exception of the Random Tree approach, were effective in achieving a classification rate that was greater than 80%. The J48 and RFT algorithms achieved the highest classification accuracy of 84.75%.

Other studies revealed that variables related to the respondent profile had very little influence on the respondents' shopping intention. In order to propose a classifier that can anticipate the purchase intention based on the behavior of users on e-commerce websites, Vieira (2015) conducted a research that compared standard machine learning techniques. The naive BaYa classifier, C4.5 decision tree algorithm, and RFT were the three methods Baati & Mohsil (2020) utilized to forecast the shopping intents of website visitors immediately after they visited a website. The results confirmed that the RFT outperformed other algorithms when it came to accuracy. There are only few articles in the current literature discussing the use of the C5.0 and CART algorithms to predict the purchasing frequency of online shopping consumers.

This present study employed the C5.0 algorithm for its capability to fix problems, such as missing values and large data. It is also capable of fast retraining data for purposes of data testing. The C5.0 algorithm features a boosting mechanism that can improve accuracy. This technique has the potential to prune decision trees, which are used to minimize over-fitting issues caused by the presence of irrelevant data. Further, in comparison to alternative classification methods, the CART algorithm exhibits several notable advantages, including enhanced interpretability of results, increased accuracy, and faster processing efficiency. Additionally, the CART algorithm can be employed for data sets with huge numbers, an array of variables, and mixed-variable scales using the binary sorting technique (Lewis, 2000; Lin & Fan, 2019). Building an optimal classification tree requires the use of training data, whereas testing the data is used to validate the model's predicting capability on new and unknown data.

### **3. RESEARCH METHODS**

#### **3.1. Decision Tree Methodology**

The process of obtaining useful information from the massive data sets was referred to as 'data mining', and it employed a number of techniques and algorithms. In the process of generating decision trees, many alternative algorithms, such as the CART, C4.5, and C5.0, could be utilized. The spread of classification approaches based on the decision trees had developed concurrently with the advances of scientific knowledge. The data in the decision tree was typically expressed in the form of a table with attributes and records. The decision tree had a clear hierarchical structure that was easy to understand. A parameter known as an attribute served as a criterion in the development of a tree (Yamasari, Nugroho, Yoshimoto, Takahashi, & Purnomo, 2019). The decision tree was comprised of a number of algorithms that collaborated to construct the tree and identify the characteristic that led to the most accurate data classification (Huynh-Cam, Chen, & Le, 2021). A range of different measures could be used to construct the classification criteria. Each decision tree algorithm applied its own unique measure to choose between the various attributes at each phase of the development process. According to Šebalj *et al.* (2017), the decision trees used the data mining methods in the form of a recursive structure to solve classification and prediction

problems by using nodes or segments as inputs to the algorithm. For this reason, an attribute that would be placed on the root node must be selected, and one branch for each potential value must be created. This created multiple subsets within the instance set, one for each of the attribute values. The operation could be performed recursively on each branch. If two decision tree nodes had the same classification, then one of the branches would be removed.

In this study, the C5.0 and CART algorithms, which were examples of data mining method for classifiers, were utilized. The purpose of using a classification tree was to classify the available data into manageable classifications that were consistent with one another. In this case, the homology revealed that the split nodes were purer, implying that each class had a substantial representation within each of node (Wu, Yang, Wu, & Han, 2019). The algorithms were employed for a variety of reasons, including dealing with massive data and performing analysis in a more timely manner. It was also possible to eliminate a number of attributes that had no significant impact on the class classification. The data of this study was collected through a structured questionnaire of online shopping consumers in Indonesia. There was a total of 820 respondents participating in this study. The structured questionnaire included the respondents' demographic information, and questionnaire items on the determinants of the online shopping consumers' purchase intentions and types of items purchased.

**3.2. C5.0 Algorithm**

The C5.0 algorithm was one of the data mining classification techniques (Giustin, Sari, & Padilah, 2022). The grouping of data in the data mining techniques was determined by the relationship between the data and sample with a label or target class. The C5.0 algorithm was one of the specialized classification algorithms available in the decision trees. It was created by improving on the ID3 and C4.5 algorithms introduced by Ross Quinlan in 1987 (Assiroj, Warnars, & Fauzi, 2018). The classification methods had been expanded into the C5.0 algorithm, designed to deal with the big data sets. The attributes were selected based on the information collecting process.

Further, the C5.0 algorithm generated a tree with varying number of branches per node. This algorithm treated continuous variables the same as the CART did. However, for categorical variables, the C5.0 algorithm treated the values of categorical variables as a splitter. Following the branching, the sample subset would be split again. The process would be repeated until the sample subset could no longer be shared. In the end, the sample subset that did not provide a major contribution to the model would be excluded.

In addition, the C5.0 algorithm generated information in the form of a decision tree or rule. The decision tree was formed based on the generated nodes. The decision tree nodes were selected depending on the gain value and entropy (Sowmya & Suneetha, 2017). The C5.0 algorithm had been widely used in classification researches (Gupta, Uttarakhand, & Rawat, 2017). The option with the most information gain would be chosen. The following is the formula representing the C5.0 algorithm:

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m p_i * \log_2(p_i) \dots \dots \dots (1)$$

Note:

S = Case

$S_m$  = Total sample

$p_i$  = Class proportion

The following formula was utilized to obtain the subset value information:

$$E(A) = \sum_{j=1}^y \frac{S_{1j} + \dots + S_{mj}}{S} I(S_{1j}, \dots, S_{mj}) \dots \dots \dots (2)$$

Note:

$$\frac{S_{1j} + \dots + S_{mj}}{S} = \text{Total subsets } j \text{ divided by the total samples.}$$

The following formula was utilized to obtain the gain value:

$$\text{Gain}(A) = I(S_1, S_2, \dots, S_m) \dots \dots \dots (3)$$

Note:

A = Attribute

S = Cases

$S_m$  = Total sample

Classifying the data with the C5.0 algorithm generated decision models and rules. These decision rules would be applied to select the best option. The C5.0 algorithm allowed a greater level of clarity, rapidly made decisions, and significantly lower memory usage, compared to the previous algorithms (Benediktus & Oetama, 2020).

### 3.3. CART Algorithm

The CART algorithm was such a widely employed predictive model in the field of machine learning. It aimed to uncover the relationships between the target variable and a number of other factors, enabling the prediction of the target variable's values. The decision tree was structured when each branch represented a predictor variable, and each node provided a prediction for the target variable upon reaching its conclusion.

Furthermore, it was also a combination of classification and regression trees (CART), which utilized the values from a collection of predictor variables in order to split the data set into homogenous subgroups that were categorically or continuously dependent on one another (Chen, Sain, & Guo, 2012; Lee, Jung, Lee, Kim, & Park, 2021). The homogeneous groups referred to the groups that all belonged to the same class in terms of the programming process. In the continuous example, the homogeneous groups were the ones with the same values in the outcomes – the values that were very similar to the mean. This study constructed the decision tree using the R Studio software's built-in machine learning algorithm.

Using the threshold value related to a specific attribute, the decision tree algorithm divided the nodes into sub-nodes. The training set was initially designated as the root node, which was then divided into two separate nodes based on the optimal attribute and threshold value. Moreover, the subsets were divided using the identical rationale. This process continued until either the last pure subset was identified or the maximum number of leaves attainable in the expanding tree had been reached. The CART algorithm was utilized to classify the data for its convenience in interpreting

and presenting the data, minimum effort for data preparation and inexistence influence caused by the nonlinearity between variables in the dataset (Zacharis, 2018).

### 3.4. Research Framework

This study employed two data mining classification algorithms, namely the C5.0 and CART. They were used to classify and predict the online shopping consumers' purchase intention, and identify several rules from the large database. The following Figure 1 presents the research framework of this study.

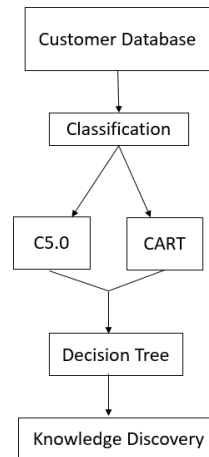


Figure 1. Research Framework

### 3.5. Data

The data was collected using a structured questionnaire. This study involved 820 respondents. The respondents must be between 18-45 years old and had purchased products through the e-commerce platforms in the previous month. This study found that most of the respondents came from West Java, East Java, and Jakarta. Several enumerators were assigned to collect the data by visiting the respondents and directly recording the data online through Google Forms. The respondents responded to a 'yes' or 'no' answer to the questionnaire items. The nominal scale was used as a measure. The first section of the questionnaire elaborated the determinants of online shopping decisions, consisting of topics of product origin, product type, and considerations for purchasing the product and shopping online through the e-commerce platforms. The price, quality, usability, need, clarity of product information, brands, and products were the elements that must be considered in purchasing a product. The respondents summarized several reasons for shopping online, which included time saving, direct home delivery, no need to travel, lower price, unlimited shop operating hours, better product quality, and better service. This study began with ethical clearance before collecting the data. The purposive random sampling method was employed.

## 4. DATA ANALYSIS AND DISCUSSIONS

### 4.1. Descriptive Statistics

This study used a structured questionnaire which was completed by 818 respondents for further analysis. Based on the results of this study, most of the respondents were female (59.02%),

between 20-25 years old (33.17%), had graduated high school and got into university (x%), and were still students (13.05%).

Furthermore, the dependent variable of this study is the amount of times spent for online shopping during the course of the previous month. The frequency of making purchases online was divided into two categories: frequently and occasionally. The “frequently” category referred to once a week, once every two to three weeks, and once a month at the very least. Meanwhile, the “occasionally” category referred to somewhere between once every two to six months and once a year. The following Table 1 presents the frequency of online shopping by the respondents of this study.

**Table 1. Frequency of Online Shopping**

Category	Frequency	Percentage
Frequently	636	77.5%
Occasionally	184	22.5%

Source: Processed data (2020)

According to Ente, Thamrin, Kuswanto, Arifin, & Andreza (2020), the chance of overfitting increased as more attributes were included in the data. Thus, an attribute selection test was performed. The Chi-Square test was performed as part of the attribute selection process, aiming to identify the predictor characteristics influencing the response characteristic. The test was applied to the attributes, and the results were quantified using the p-value. Comparison of the p-value to a significant level of 5% was used to determine the best attributes to be selected. The following Table 2 presents the results of Chi-Square test which initially involved 72 attributes, but the final stage only comprised 38 explanatory attributes as follows:

**Table 2. Results of Chi-Square Test**

Attribute (x)	p-Value	Conclusion	Attribute (x)	p-Value	Conclusion
<b>Respondent Profile</b>			<b>Product Type</b>		
Job position	0.013	Significant	Electronic accessories	0.000	Significant
Occupation	0	Significant	Electronic equipment	0.000	Significant
<b>E-commerce Platform</b>			Men fashion	0.001	Significant
Shopee	0.000	Significant	Women fashion	0.000	Significant
Tokopedia	0.000	Significant	Beauty	0.001	Significant
Bukalapak	0.000	Significant	Fashion accessories	0.000	Significant
Lazada	0.000	Significant	Furniture and Life Style	0.000	Significant
Social Media	0.000	Significant	Home electronics	0.000	Significant
JD.ID	0.000	Significant	Groceries	0.000	Significant
Blibli	0.000	Significant	Fast food	0.000	Significant
<b>Reasons to Shop Online</b>			Health product	0,000	Significant
Time saving	0.000	Significant	Toys	0.000	Significant
Direct home delivery	0.000	Significant	Mom and baby product	0.000	Significant
Wide range of products	0.000	Significant	Sports and outdoor	0.000	Significant
No need to travel	0.000	Significant	Automotive	0.000	Significant
Lower price	0.000	Significant	<b>Buying Priority</b>		
Unlimited shop operating hours	0.000	Significant	Quality	0.009	Significant
Good quality	0.000	Significant	Negotiable product	0.000	Significant

Source: Processed data (2020)



#### 4.2. Results of C5.0 Algorithm

The following Table 3 presents the results of C5.0 algorithm, which confirms that the attributes such as occupation, social media (Instagram and Facebook), product origin, and product quality all contribute 100% to the purchase decision. Further, Shopee as one of the e-commerce platforms has a quite high contribution of 99.37%. Meanwhile, the mother and baby product and fast food categories also have a quite high contribution of 98.62% and 92.60%, respectively. Additionally, time saving was the most common reason for shopping online (96.86%).

**Table 3. Results of C5.0 algorithm**

Attribute	Percentage
Occupation	100.00%
Social media	100.00%
Product origin	100.00%
Product quality	100.00%
Shopee	99.37%
Electronic accessories	98.62%
Time saving	96.86%
Fast food	92.60%
Blibli	91.59%
Home electronics	88.08%
Sports and outdoor	87.83%
JD.ID	84.94%
Lower price	81.05%
Mom and baby product	76.29%
Negotiable product	73.15%
Fashion accessories	72.15%
Profession	68.13%
Tokopedia	67.13%
Furniture and life style	67.13%
Toys	64.74%
Wide range of products	60.23%

Source: Processed data (2020)

There were several conditions that could predict the consumers' online shopping intention based on the frequency of online shopping. This study categorized the frequency of online shopping to "frequently" and "occasionally" categories, according to the C5.0 algorithm classification tree. The results of this study confirm that the consumers would likely be in the "frequently" category when they were (1) employed, especially as supervisors or entrepreneurs; (2) employed and had purchased the products online through the social media platforms; and (3) had purchased the products online while paying attention to the product quality. These indicated that the consumers' occupation was the most influential determinant of the consumers' online shopping intention and behavior. Income was highly influential to the consumers' socio-economic class and the frequency of online shopping. These findings are consistent with those of Pallabi (2015), Rudansky-Kloppers (2014), Vasic *et al.* (2019), and Venkatesh, Speier-Pero, & Schuetz (2022) who argued that the employment and income were predictors of online shopping intention.

Individuals with high income frequently shopped online, whereas those with lower income required more time to make decisions and purchase when necessary.

In addition, factors such as employment characteristics, social media platforms, product quality, and country of origin were also the most accurate predictors of online shopping behavior. According to the findings of the C5.0 algorithm, the consumers who regularly used the social media platforms had a significant advantage when they shopped online. Similarly, Zhang, Trusov, Stephen, & Jamal (2017) found that there was a positive relationship between the cumulative use of social networks over time and shopping behaviors. Many companies had recognized the potential of social media in expanding their consumer base and chosen to capitalize on these opportunities. In addition, people used the social media platforms for a variety of purposes, including entertainment, professional networks, and learning about new products. The consumers were able to make purchase decisions based on the reviews and comments found available on the social media platforms. At the moment, the social media platforms served as a communication environment between buyers and sellers. In this setting, the buyers and sellers had the opportunity to develop relationships with one another in addition to making transactions (Jothi & Gaffoor, 2017). The growing involvement of consumers – as the buyers – and the sellers demonstrated the interactive nature of social media platforms. The consumers were actively involved in the production of content, spread of information, and promotion of products through the interactions on the social media platforms. The increased accessibility the social media platforms provided to the businesses and consumers contributed to an improvement in customer service. In addition, the consumers primarily relied on the social media platforms to collect information regarding products and services to ensure that the seller profile had the useful information which impressed the consumers.

Furthermore, this study proves that the product quality is one of the determinants of online shopping intention. This suggested that the consumers were aware of their needs. The initial browsing stage in the online shopping process resulted in several considerations when making the purchase decisions. The consumers continued to focus on the quality of the product before making a purchase decision. This included determining whether the product matched the specifications and real photos. According to Wang & Le (2016) the quality of products and services was the most important factor, as it had a direct impact on the consumers' levels of satisfaction and it could shape their positive behavioral intentions. The consumers liked to shop at online shops which provided the high quality products. After finding online shops with the high quality products which even exceeding their expectations, they would likely to continue shopping at the shops frequently. In addition, in the food online shops, the consumers tended to be interested in learning more about the food they liked and would likely to repeat purchases. When they were aware of the products consumed and their origin, they could make comparisons between domestically produced goods and those imported from other countries, and even more accurately anticipate the delivery cost. The likelihood that a consumer purchasing online was equal to the decrease in the delivery cost in terms of purchasing the high-quality products.

In addition, the product type was also a determinant in the selection of products to purchase. The mom and baby products, fast food, home electronics, sports and outdoors, and electronic accessories were the type of products that were mostly purchased and contributed to the consumers' online shopping intention. This allowed for the development of product type segmentation to develop, which was one of the primary factors of online shopping.

### 4.3. Results of CART Algorithm

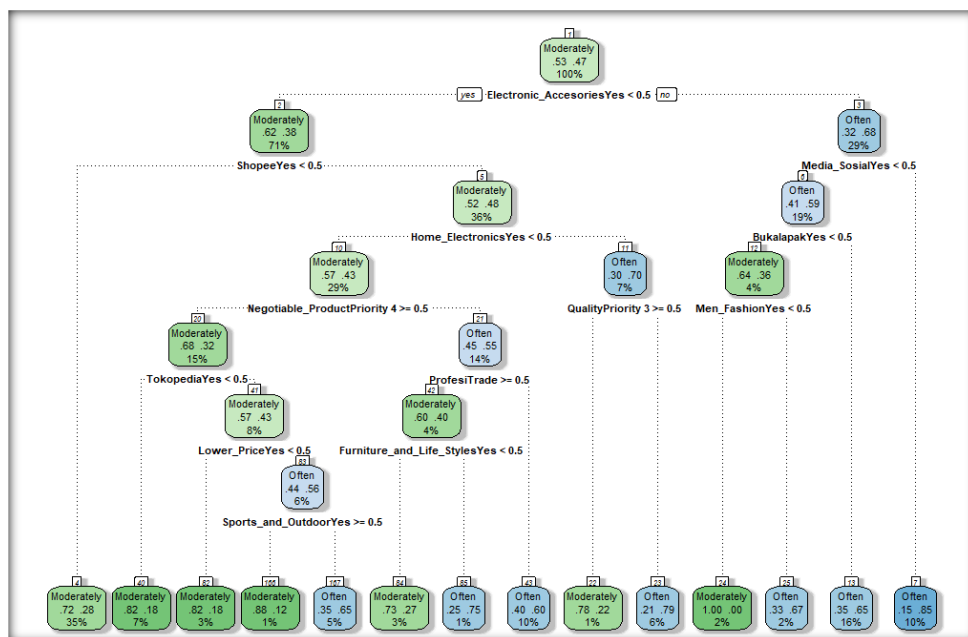
The initial analysis of the C5.0 algorithm in creating a classification tree aimed to identify the nodes serving as the tree's root, after which the analysis continued to identifying the branches associated with each node. In addition, the calculation of entropy, gain, and gain ratio, and class division was performed on the branches, and the process was repeated until every branch had a class with high data qualities that contributed to the online shopping decisions. The final stage of analysis involved the formation of a classification tree (Figure 2). The results of CART algorithm can be seen in the following Table 4.

**Table 4. Results of CART algorithm**

Attribute	Percentage
Job position	100.00%
Social media	100.00%
Product origin	100.00%
Product quality	100.00%
Mom and baby product	98.62%
Time saving	96.86%
Fast food	92.60%
Blibli	91.59%

Source: Processed data (2020)

Table 4 shows that the employment characteristics, social media, product origin, and product quality all have a massive influence (100%) on the consumers' purchase decision. These results are comparable to the results of C5.0 algorithm. The following Figure 2 presents a number of determinants of Indonesian consumers' purchase intention (frequently or occasionally), including the product type (electronic accessories, home electronics, men fashion, and furniture and life style), e-commerce platforms (Shopee, social media, Bukalapak, and Tokopedia), and major considerations of online shopping (negotiable product, quality and lower price).



Source: Processed data (2020)

Figure 2. CART Classification Tree

Figure 2 shows that there are several conditions that could predict the consumers' purchase intention based on the CART classification tree, as indicated by the frequency of online shopping. The following conditions picturize the consumers of "occasionally" category:

1. The consumers of electronic accessories at Shopee, with a shopping probability of 72.10%.
2. The consumers of home electronics and electronic accessories who prioritized lower and negotiable price and do not make a purchase from Tokopedia, with a probability of 81.82%.
3. The consumers of electronic accessories who do not purchase at Shopee nor Tokopedia, and instead purchase negotiable home electronics, and love sports and outdoor products, with a probability of 87.50%.
4. The consumers of electronic accessories who do not purchase them from Shopee and still consider the product quality, with a probability of 77.78% .
5. The consumers of men fashion who not buy the electronic accessories but instead purchases other products on the social media platforms, such as Bukalapak, with a probability of 100.00%.
6. The consumers of electronic accessories who do not purchase at Shopee, purchase the negotiable home electronics, work at the trade sector, and purchase products of furniture and life styles, with a probability of 72.73%,.

Meanwhile, the conditions picturizing the consumers of "frequently" category are as follow:

1. The consumers of electronic accessories who do not purchase at Shopee nor Tokopedia but purchase the negotiable home electronics, with a probability of 64.86%.
2. The consumers of electronic accessories who do not purchase at Shopee, but purchase the negotiable home electronics, work at the trade sector, and do not purchase the products of furniture and life style, with a probability of 75.00%.
3. The consumers of electronic accessories who do not purchase at Shopee, but purchase the negotiable home electronics, and do not work at the trade sector, with probability of 60.26%.
4. The consumers of electronic accessories who do not purchase at Shopee, do not purchase the home electronics, and consider the product quality, with a probability of 79.17%.
5. The consumers who do not purchase the electronic accessories but instead other products on the social media platforms and Bukalapak, with a probability of 66.67%.
6. The consumers who do not purchase the electronic accessories and purchase products at the social media platforms rather than at Bukalapak, with a probability of 64.80%.
7. The consumers who do not purchase the electronic accessories nor purchase products at the social media platforms, with a probability of 85.18%.

According to the CART algorithm rules, there were four attributes to identify the determinants of consumers' online shopping intention: product type, occupation, e-commerce platform, product quality, and low price. The consumers who occasionally shopped online frequently considered several factors before deciding to purchase the products online, including lower and negotiable prices, and product quality. They usually shopped at the social media platforms and Bukalapak, and some still shopped at Shopee. Meanwhile, the consumers who

frequently shopped online usually preferred Shopee, Tokopedia, Bukalapak, Instagram and Facebook). They also considered factors such as lower price and product quality. The electronic accessories were the most desired items.

#### 4.4. Comparison of C5.0 and CART Algorithm Models

The following Table 5 presents the results of model comparison, between the C5.0 and CART algorithms. The accuracy recorded in the following table was the result of the data iteration process and the previously created confusion matrix. The C5.0 and CART algorithms predicted the Indonesian consumers' online shopping intention with an accuracy of 65.51% and 63.13%, respectively.

**Table 5. Model Comparison**

C5.0		CART	
Accuracy	Kappa	Accuracy	Kappa
65.51	30.79	63.13	25.87

Source: Processed data (2020)

#### 4.5. Discussion

The analysis of C5.0 and CART algorithms indicate that the Indonesian consumers' online shopping intention was based on the occupation, social media platforms, product quality, and product origin. The product type and compatibility were substantial indicators of online shopping behavior, which is another significant findings of this present study. The product type selected was in conjunction with the accessibility of information and it tended to influence the online buying behavior and led to a new segmentation in the online shopping (Venkatesh *et al.*, 2022). Consequently, the consumers were more open to purchasing new products, which in turn led to them having more options and more knowledge about products on the market.

Further, this study also confirms that time saving was the primary reason the consumers preferred the online shopping. According to Vasic *et al.* (2019) the time was the most valuable resource that the consumers spent to consider whether to purchase the products online or offline. It was because exploring the online catalogs while shopping online saved more time and lowered their stress level compared to conventional shopping. The convenience of making purchases and the effectiveness of the use of time when shopping were found as two primary benefits of shopping online. Online transactions could be make at any location and at any time, so that the online shop consumers were less likely to experience time pressure or time limit when they made purchases. The consumers felt highly convenient as they had unlimited time, did not need to travel, did not experience traffic, and did not need to queue just like in the crowded offline stores (Rudansky-Kloppers, 2014).

In addition, the consumers paid attention to the price – where they loved low and negotiable price – and product quality. The online shops also offered interesting price compared to the offline stores. Discounts on the products at the online shops affected the consumers' confidence in the product price, which would in turn affected their level of satisfaction. A more significant effect was determined by how the price was perceived. Pricing had a direct impact on how the consumers perceived the value of a product or service, as well as on the transactions and level of customer satisfaction. A negative perception of price would negatively affect the customer satisfaction, and vice versa. The price and product quality were such two ultimate determinants of customer satisfaction, since the consumers would prefer relatively low cost for the products of high

quality, regardless of whether they made the purchases online or in-store. Satisfied consumers tended to have higher intentions to make repeat purchases and recommend the products to their acquaintances. This was in contrast to the unsatisfied consumers who tended to look for alternatives. This resulted to a higher chance of the consumers to switch to other sellers or even e-commerce platforms. Thus, the customer satisfaction could be achieved when the products exceeded the consumers' expectations and the products' perceived performance were consistent with one another.

Another significant finding of this study reveal that the e-commerce platforms were such a key node in the classification of online shopping consumers. Perceived benefits, perceived ease of use, and reported enjoyment of the online shopping intention were all inextricably related to the choice of e-commerce platforms. Shopee, social media, Bukalapak, and Tokopedia each had a significant number of nodes in the decision trees. The online shopping consumers were highly familiar of these e-commerce platforms. This suggested that the ability to attract the customers to the online shops might be based on the ease of use and convenience offered by the technology. The ease of use and overall convenience of the purchasing experience were two main selling aspects of shopping online. It was critical for the sellers and business participating in the e-commerce since the industry was growing significantly and highly profitable. Recognizing this trend allowed them to realize the importance of providing the consumers with reliable information and services, as well as appealing product content presented in a format appropriate for each country.

Another concern in the context of online shopping was that the fundamental goals of online shopping consumers were to receive high-quality products, simple service from the sellers and e-commerce platforms, and dependable delivery. In terms of delivery, the consumers were concerned about the product delivery that must be timely, safe, and dependable to the assigned address of destination. The timely delivery was critical to assure the customer satisfaction. The consumers could easily switch from one platform to another if the product quality and delivery were not satisfactory. As a result, despite the growth of online business in this modern era, business strategies must still be the focus to increase the customer satisfaction. The strategies included the pricing, product quality, and the provision of services. The satisfied consumers were considered as a vital asset for both the sellers and businesses, as well as for the business sustainability.

## **5. CONCLUSIONS, SUGGESTIONS, AND LIMITATIONS**

The results of this study have a number of implications. It is confirmed that the product type, e-commerce platform, product quality, and low price of products are the most significant determinants for the classification of consumers' online shopping frequency. The consumers' occupation, and variables related to their profile have essentially no effect on the consumers' online shopping intention. The considerations for purchasing the products online include the online shopping experience, where the less uncertain they feel about the product quality and the benefits, the less risk they will perceive when making online purchases.

The decision trees constructed using the C5.0 and CART algorithms are able to accurately classify the data into both of the categories. All of the models have a categorization rate higher than 60%. The C.50 algorithm has the highest classification accuracy, reaching 65.51%. Considering that there is no statistically significant difference between the two algorithms, this

study employed the C.50 algorithm for its excellent capabilities. The decision trees constructed performed just as well as expected when it came to classifying the two groups. In relation to the online shopping frequency classification, the most important factors to consider are the product type, e-commerce platforms, product quality, and affordable price. The consumers' occupation, and variables related to their profile have essentially no effect on the consumers' online shopping intention. However, this study has faced several limitations. The categorization model of this study was based on observational studies and involved many variables to identify the determinants of online shopping intention. Future studies are suggested to use different variables, and sets of rules. In addition, future researches are suggested to use other algorithms to compare and obtain the best classification accuracy.

## REFERENCES

- Ali, F., & Sohail, M. (2018). Effects of Corporate Social Responsibility on Consumer Purchase Intention. *Pakistan Journal of Humanities and Social Sciences*, 6(4), 477–491. [www.pjhss.com](http://www.pjhss.com)
- Aliedan, M. M., Elshaer, I. A., Alyahya, M. A., & Sobaih, A. E. E. (2022). Influences of University Education Support on Entrepreneurship Orientation and Entrepreneurship Intention: Application of Theory of Planned Behavior. *Sustainability (Switzerland)*, 14(20). <https://doi.org/10.3390/su142013097>
- Alsoud, M. A. S., & Othman, I. bin L. (2018). The Determinant of Online Shopping Intention in Jordan: A Review and Suggestions for Future Research. *International Journal of Academic Research in Business and Social Sciences*, 8(8). <https://doi.org/10.6007/ijarbss/v8-i8/4507>
- Assiroj, P., Warnars, H. L. H. S., & Fauzi, A. (2018). Comparing CART and C5.0 Algorithm Performance of Human Development Index. *2018 Third International Conference on Informatics and Computing (ICIC)*, 1–5. <https://doi.org/10.1109/IAC.2018.8780439>
- Baati, K., & Mohsil, M. (2020). Real-Time Prediction of Online Shoppers' Purchasing Intention Using Random Forest. In *Artificial Intelligence Applications and Innovations* (pp. 43–51). [https://doi.org/10.1007/978-3-030-49161-1\\_4](https://doi.org/10.1007/978-3-030-49161-1_4)
- Beck, S., & Kenning, P. (2015). The Influence of Retailers' Family Firm Image on New Product Acceptance: An Empirical Investigation in the German FMCG Market. *International Journal of Retail & Distribution Management*, 43(12), 1126–1143. <https://doi.org/10.1108/IJRDM-06-2014-0079>
- Benediktus, N., & Oetama, R. S. (2020). Algoritma Klasifikasi Decision Tree C5.0 untuk Memprediksi Performa Akademik Siswa. *14 ULTIMATICS*, XII(1). <https://www.kaggle.com/aljarah/xAPI-Edu-Data>
- Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing and Customer Strategy Management*, 19(3), 197–208. <https://doi.org/10.1057/dbm.2012.17>
- Chua, T. H. H., & Chang, L. (2016). Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. *Computers in Human Behavior*, 55, 190–197. <https://doi.org/10.1016/j.chb.2015.09.011>

- Ente, D. R., Thamrin, S. A., Kuswanto, H., Arifin, S., & Andreza. (2020). Klasifikasi Faktor-Faktor Penyebab Penyakit Diabetes Melitus di Rumah Sakit UNHAS Menggunakan Algoritma C4.5 \*. *Indonesian Journal of Statistics and Its Applications*, 4(1), 80–88.
- Giningroem, D. S. W. P., Setyawati, N. W., & Wijayanti, M. (2022). Consumer Experiences, Time Saving Orientation, and Price Saving Orientation on Actual Behavior to Use Application Online Food Delivery through Convenience Motivation. *East Asian Journal of Multidisciplinary Research*, 1(11), 2549–2560. <https://doi.org/10.55927/eajmr.v1i11.1989>
- Giustin, F. N., Sari, B. N., & Padilah, T. N. (2022). Application of C5.0 Algorithm in Prediction of Learning Outcomes in Calculus Subject. In *Journal of Applied Engineering and Technological Science* (Vol. 3, Issue 2).
- Gopinath, R. (2019). Online Shopping Consumer Behaviour of Perambalur District. *International Journal of Research*, VIII(V), 542–547. <https://www.researchgate.net/publication/332961054>
- Gu, S., & Wu, Y. (2019). Using the Theory of Planned Behaviour to Explain Customers' Online Purchase Intention. *World Scientific Research Journal*, 5. [https://doi.org/10.6911/WSRJ.201909\\_5\(9\).0026](https://doi.org/10.6911/WSRJ.201909_5(9).0026)
- Gupta, B., Uttarakhand, P., & Rawat, I. A. (2017). Analysis of Various Decision Tree Algorithms for Classification in Data Mining. In *International Journal of Computer Applications* (Vol. 163, Issue 8).
- Ha, N. T., Nguyen, T. L. H., Pham, T. Van, & Nguyen, T. H. T. (2021). Factors Influencing Online Shopping Intention: An Empirical Study in Vietnam. *Journal of Asian Finance, Economics and Business*, 8(3), 1257–1266. <https://doi.org/10.13106/jafeb.2021.vol8.no3.1257>
- Huynh-Cam, T. T., Chen, L. S., & Le, H. (2021). Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance. *Algorithms*, 14(11). <https://doi.org/10.3390/a14110318>
- Jothi, C. A., & Gaffoor, A. M. (2017). Impact of Social Media in Online Shopping. *ICTACT Journal on Management Studies*, 3(3), 576–586. <https://doi.org/10.21917/ijms.2017.0079>
- Kardinasari, R., Iskandar, T. Z., Nugraha, Y., & Jatnika, R. (2019). Social Sensitivity Effect to Public Service Competence and Its Impact on the Head of Sub-district Performance in West Java Province. *Journal of Psychology Research*, 9(1). <https://doi.org/10.17265/2159-5542/2019.01.004>
- Karimi, S., Holland, C. P., & Papamichail, K. N. (2018). The Impact of Consumer Archetypes on Online Purchase Decision-Making Processes and Outcomes: A Behavioural Process Perspective. *Journal of Business Research*, 91, 71–82. <https://doi.org/10.1016/j.jbusres.2018.05.038>
- Lee, J. E., & Chen-Yu, J. H. (2018). Effects of price discount on consumers' perceptions of savings, quality, and value for apparel products: mediating effect of price discount affect. *Fashion and Textiles*, 5(1). <https://doi.org/10.1186/s40691-018-0128-2>
- Lee, J., Jung, O., Lee, Y., Kim, O., & Park, C. (2021). A comparison and interpretation of machine learning algorithm for the prediction of online purchase conversion. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(5). <https://doi.org/10.3390/jtaer16050083>
- Lewis, R. J. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis*. <https://www.researchgate.net/publication/240719582>



- Lim, Y. J., Osman, A., Salahuddin, S. N., Romle, A. R., & Abdullah, S. (2016). Factors Influencing Online Shopping Behavior: The Mediating Role of Purchase Intention. *Procedia Economics and Finance*, 35, 401–410. [https://doi.org/10.1016/S2212-5671\(16\)00050-2](https://doi.org/10.1016/S2212-5671(16)00050-2)
- Lin, C. L., & Fan, C. L. (2019). Evaluation of CART, CHAID, and QUEST algorithms: a case study of construction defects in Taiwan. *Journal of Asian Architecture and Building Engineering*, 18(6), 539–553. <https://doi.org/10.1080/13467581.2019.1696203>
- Lu, L.-C., Chang, W.-P., & Chang, H.-H. (2014). Consumer Attitudes toward Blogger's Sponsored Recommendations and Purchase Intention: The Effect of Sponsorship Type, Product Type, and Brand Awareness. *Computers in Human Behavior*, 34, 258–266. <https://doi.org/10.1016/j.chb.2014.02.007>
- Meeprom, S., & Silanoi, T. (2020). Investigating the perceived quality of a special event and its influence on perceived value and behavioural intentions in a special event in Thailand. *International Journal of Event and Festival Management*, 11(3), 337–355. <https://doi.org/10.1108/IJEFM-09-2019-0043>
- Pallabi, M. (2015). Motivator of Online Shopping: The Income Factor. *Asian Journal of Research in Banking and Finance*, 5(11), 34–46. <https://doi.org/10.5958/2249-7323.2015.00132.7>
- Pasharibu, Y., Soerijanto, J. A., & Jie, F. (2020). Intention to Buy, Interactive Marketing, and Online Purchase Decisions. *Jurnal Ekonomi Dan Bisnis*, 23(2), 339–356.
- Prasad, S., Garg, A., & Prasad, S. (2019). Purchase decision of generation Y in an online environment. *Marketing Intelligence and Planning*, 37(4), 372–385. <https://doi.org/10.1108/MIP-02-2018-0070>
- Rehman, A. U., Bashir, S., Mahmood, A., Karim, H., & Nawaz, Z. (2022). Does e-shopping service quality enhance customers' e-shopping adoption? An extended perspective of unified theory of acceptance and use of technology. *PLoS ONE*, 17(2 February). <https://doi.org/10.1371/journal.pone.0263652>
- Rehman, S. F. U., & Ansari, M. F. (2023). Predicting Customer Satisfaction of Online Shoppers Using AI – A Theoretic Framework. *IJARCCCE*, 12(1). <https://doi.org/10.17148/ijarccce.2023.12112>
- Rudansky-Kloppers, S. (2014). Investigating Factors Influencing Customer Online Buying Satisfaction In Gauteng, South Africa. *International Business & Economics Research Journal*, 13(5), 1187–1198.
- Saleem, A., Aslam, J., Kim, Y. B., Nauman, S., & Khan, N. T. (2022). Motives towards e-Shopping Adoption among Pakistani Consumers: An Application of the Technology Acceptance Model and Theory of Reasoned Action. *Sustainability (Switzerland)*, 14(7). <https://doi.org/10.3390/su14074180>
- Šebalj, D., Franjković, J., & Hodak, K. (2017). Shopping Intention Prediction using Decision Trees. *Millenium - Journal of Education, Technologies, and Health*, 2(4), 13–22. <https://doi.org/10.29352/mill0204.01.00155>
- Sinha, J., & Kim, J. (2012). Factors Affecting Indian Consumers' Online Buying Behavior". *Innovative Marketing*, 8(2), 46–57. [www.tradechakra.com](http://www.tradechakra.com),
- Sowmya, R., & Suneetha, K. R. (2017). *Data Mining with Big Data*.
- Sussman, R., & Gifford, R. (2019). Causality in the Theory of Planned Behavior. *Personality and Social Psychology Bulletin*, 45(6), 920–933. <https://doi.org/10.1177/0146167218801363>

- Truong, N. X. (2018). The Impact of Hallyu 4.0 and Social Media on Korean Products Purchase Decision of Generation C in Vietnam. *The Journal of Asian Finance, Economics and Business*, 5(3), 81–93. <https://doi.org/10.13106/jafeb.2018.vol5.no3.81>
- Vasic, N., Kilibarda, M., & Kaurin, T. (2019). The Influence of Online Shopping Determinants on Customer Satisfaction in the Serbian Market. *Journal of Theoretical and Applied Electronic Commerce Research*, 14(2), 70–89. <https://doi.org/10.4067/S0718-18762019000200107>
- Venkatesh, V., Speier-Pero, C., & Schuetz, S. (2022). Why do People Shop Online? A Comprehensive Framework of Consumers' Online Shopping Intentions and Behaviors. *Information Technology & People*, 35(5), 1590–1620. <https://doi.org/10.1108/ITP-12-2020-0867>
- Vieira, A. (2015). Predicting Online User Behaviour using Deep Learning Algorithms. *Machine Learning*, 1–21. <http://arxiv.org/abs/1511.06247>
- Vijay, T. S., Prashar, S., & Sahay, V. (2019). The Influence of Online Shopping Values and Web Atmospheric Cues on E-Loyalty: Mediating Role of E-Satisfaction. *Journal of Theoretical and Applied Electronic Commerce Research*, 14(1), 1–15. <https://doi.org/10.4067/S0718-18762019000100102>
- Vyas, A., & Bissa, G. (2017). A Study on Customer Preference towards Online Shopping with Special Reference to Bikaner City. *International Journal of Engineering Technology Science and Research (IJETSR)*, 4(9), 675–681.
- Wahyudin, Moh., Yuliando, H., & Savitri, A. (2020). Consumer Behavior Intentions to Purchase Daily Needs through Online Store Channel. *AgriTECH*, 40(4), 306–311. <https://doi.org/10.22146/agritech.49232>
- Wang, L. W., & Le, Q. L. (2016). Customer Satisfaction towards Online Shopping at Electronics Shopping Malls in Vietnam- A Conceptual Model to Enhance Business Success through Efficient Websites and Logistics Services. *Journal of Stock & Forex Trading*, 05(01). <https://doi.org/10.4172/2168-9458.1000164>
- Wu, C., Yang, F., Wu, Y., & Han, R. (2019). Prediction of Crime Tendency of High-Risk Personnel using C5.0 Decision Tree Empowered by Particle Swarm Optimization. *Mathematical Biosciences and Engineering*, 16(5), 4135–4150. <https://doi.org/10.3934/mbe.2019206>
- Yamasari, Y., Nugroho, S. M. S., Yoshimoto, K., Takahashi, H., & Purnomo, M. H. (2019). Expanding Tree-Based Classifiers using Meta-Algorithm Approach: An Application for Identifying Students' Cognitive Level. *International Journal of Innovative Computing, Information and Control*, 15(6), 2085–2107.
- Zacharis, N. Z. (2018). Classification and regression trees (CART) for predictive modeling in blended learning. *International Journal of Intelligent Systems and Applications*, 10(3), 1–9. <https://doi.org/10.5815/ijisa.2018.03.01>
- Zhang, Y., Trusov, M., Stephen, A. T., & Jamal, Z. (2017). Online Shopping and Social Media: Friends or Foes? *Journal of Marketing*, 81(6), 24–41. <https://doi.org/10.1509/jm.14.0344>

#### ADDITIONAL REFERENCES

APJII. (2022). *Profil Internet Indonesia 2022 (Issue June)*.

**APPENDIX****Items of Questionnaire**

No	Questions	Type of Answers
1	Demographics Questions	Multiple choice
1	Frequency of Online Shopping	Likert Scale
2	Platform / E-commerce	Multiple Answers
3	Reasons of online shop	Multiple Answers
4	Kind of Stuff	Multiple Answers
5	Your main consideration in choosing goods	Priority / Likert Scale
6	Challenges of online buying	Multiple Answers