Preliminary communication
UDC: 658.89:004.7(497.5)
https://doi.org/10.18045/zbefri.2023.1.157

Decision trees do not lie: Curiosities in preferences of Croatian online consumers*

Ana Marija Filipas[1], *Nenad Vretenar*[2], *Ivan Prudky*[3]

Abstract

Understanding consumers' preferences has always been important for economic theory and for business practitioners in operations management, supply chain management, marketing, etc. While preferences are often considered stable in simplified theoretical modelling, this is not the case in real-world decision-making. Therefore, it is crucial to understand consumers' preferences when a market disruption occurs. This research aims to recognise consumers' preferences with respect to online shopping after the COVID-19 outbreak hit markets. To this purpose, we conducted an empirical study among Croatian consumers with prior experience in online shopping using an online questionnaire. The questionnaire was completed by 350 respondents who met the criteria. We selected decision-tree models using the J48 algorithm to determine the influences of the found shopping factors and demographic characteristics on a consumer's preference indicator. The main components of our indicators that influence consumer behaviour are the stimulators and destimulators of online shopping and the importance of social incidence. Our results show significant differences between men and women, with men tending to use fewer variables to make decisions. In addition, the analysis revealed that four product groups and a range of shopping mode-specific influencing factors are required to evaluate consumers' purchase points when constructing the consumers' preference indicator.

Keywords: *decision-making, consumers' preferences, data mining, decision trees, shopping behaviour indicators*

JEL classification: *C44, D12*

---

* Received: 24-05-2023; accepted: 28-06-2023

[1] *Teaching assistant, University of Rijeka, Faculty of Economics and Business, Ivana Filipovića 4, 51000, Rijeka, Croatia. Scientific affiliation: quantitative economics, econometrics, statistics. Phone: +385 51 355 162. E-mail: ana.marija.filipas@efri.hr.*

[2] *Associate professor, University of Rijeka, Faculty of Economics and Business, Ivana Filipovića 4, 51000, Rijeka, Croatia. Scientific affiliation: organization theory, operational management, decision theory. Phone: +385 51 355 163. E-mail: nenad.vretenar@efri.hr.*

[3] *Teaching assistant, University of Rijeka, Faculty of Economics and Business, Ivana Filipovića 4, 51000, Rijeka, Croatia. Scientific affiliation: quantitative economics, decision theory, sports economics. Phone: +385 51 355 161. E-mail: ivan.prudky@efri.hr.*

1. Introduction

The adoption of online shopping differs between Croatia and the rest of the EU countries, as Croatia has some catching up to do, compared to these countries. Some of the reasons for the lag in adoption could be explained by understanding what is pulling Croatian consumers away from engaging in online shopping. With an average number of landline internet connections per 100 inhabitants of 35.5 in the EU-27 in 2020, Croatia is at the bottom of the list with 25.1 landline internet connections per 100 inhabitants, surpassing only Poland (Our World in Data, 2022). A lower number of internet users buying products is also evident from the number of people who bought products online in 2019: 34.97% of Croatians bought products online, while the EU-27 average is 48.97% (Eurostat, 2023-a). Interestingly, men buy more online (38.98% in Croatia and 49.81% in EU-27) than women (31.17% in Croatia and 48.15% in EU-27). However, fewer landline internet connections are not the primary determinant of Croatian consumers' shopping habits.

Various perceived barriers and problems in online shopping shape their decisions. According to Eurostat data (2023-b), individuals prefer shopping in person (29.04%) much more than the EU-27 average (17.40%), and they like to see the product while shopping, are loyal to stores, are forced to buy in stores, or do so out of habit. Another barrier to shopping online, again much higher than the EU-27 average, is concern about the cost of delivery of goods (Croatia: 7.07% of individuals; EU-27 average: 2.79% of individuals). Concern about payment security or privacy as the next main barrier is again higher in Croatia than in the EU-27 average: 5.82% and 5.61% of individuals, respectively, report such concerns. The reasons for such behaviour are also not supported by factual circumstances, as online shopping problems are less common in Croatia than in the EU-27 average (Eurostat, 2023-c). Although 31.50% of individuals did not experience problems purchasing through a website or app for personal use (4.90 PP. less than the EU-27 average), only 13.87% of individuals reported encountering problems while making online purchases (4.83 PP. less than the EU-27 average). The most important problem is the duration of delivery, which is criticised by 6.98% of individuals in Croatia, compared to an average of 11.82% of people in the EU-27. The use of websites that are too complicated or do not work satisfactorily is the second most common problem, reported by 4.10% of individuals in Croatia, which is also lower than the EU-27 average of 5.29%. A more common problem than the EU-27 average is the foreign seller not selling in their country as 3.41% of Croatian respondents struggled with this, compared to an EU-27 average of 2.28%. Other, less common problems, such as faulty or damaged delivered goods, difficulty finding information, satisfactory responses to complaints about products, or final costs that differ from those quoted, were all less common than the EU-27 averages. Different views of online shopping influence the decision to use it, but they are not the only determining factors. According to the data presented, people in Croatia perceive fewer barriers to online shopping than the EU-27 average. The problems they encountered when shopping online were also less frequent than the EU-27 average. Nevertheless, Croatians prefer shopping in stores for several reasons that should be explored.

The observed curiosities lead to the formulation of some working research hypotheses: *consumers choose their preferred shopping method differs depending on sociodemographic subgroups; shopping mode-specific factors influence consumers' shopping decisions, and men and women have different shopping decision patterns.* Mode-specific factors are undoubtedly important in the decision between shopping in stores and shopping online. However, different sociodemographic characteristics of consumers also influence the final decision on how to shop. Shifts in influencing factors can be observed especially during and after the market disruption of COVID-19. The factors influencing consumers' decisions on their preferred mode of purchase are the subject of this paper, considering this decision as the dependent variable. The manner and the importance of the influence of the shopping mode-specific characteristics, as well as the sociodemographic characteristics of the consumers, are considered variables that influence the consumers' final decision on the preferred shopping types.

The *Literature review* section of this paper presents previous findings on consumers' preferences and the theoretical grounds for the chosen methods. A more detailed insight into the use of algorithms, tests, and constructed (in)dependent variables is provided in the *Methodology* section. The *Empirical Data and Analysis* section provides a general overview of the data collection, the research sample and a detailed insight into the analysis conducted, which leads directly into the *Results and Discussion section*, where the main findings are compared to previous ones and the new insights and main contributions of our research are highlighted. The *Conclusion* section highlights the main findings and suggests possible lanes for future research.

2. Literature review

A variety of factors influence consumers' shopping preferences. Consumers' preferences can change fundamentally when hit by severe market disruptions, both short- and long-term. Thus, the influencing factors to a consumer's shopping decision will be looked at from two separate aspects: the *sociodemographic characteristics* of consumers and *shopping mode specifics* as influencing variables. The sociodemographic characteristics found to be most influential in previous research were gender, age, employment or work status, education level, personal income, and urbanisation level of the residence. Regarding shopping mode-specific factors, they proved to have a positive or stimulative, or negative or destimulative influence on consumers' shopping decisions.

Gender differences are noticeable when choosing the shopping mode (Kim et al., 2020). According to Yahya and Sugiyanto (2020), who studied the effects of demographic and socioeconomic factors on a person's decision to shop online, women are more likely to shop online than men. Further, consistent conclusions were presented in the research papers of Hood et al. (2020) and Moon et al. (2021), demonstrating a higher likeliness of women to shop online than men. Moreover, women will spend more shopping online than men (Truong and Truong, 2022). Undesirable aspects such as prolonged delivery time don't affect women's decision to shop online or not. Even when delivery times are longer, women are more likely to shop online than men (Dias et al., 2022).

Expected is that distinct *age groups* have different attitudes towards shopping modes. Hood et al. (2020) found statistically significant evidence that people aged 25 to 44 are more likely to shop online than other age groups. Similar results came from Moon et al. (2021), stating consumers in their late 20s and 30s are most likely to shop online. Dias et al. (2022) and Buhaljoti et al. (2022) discovered that consumers' intentions to shop online decrease as they age. Giannakopoulou et al. (2022) conducted a study in Cyprus and found that age influences online grocery shopping. Younger individuals, in particular, tend to shop online more frequently. Rummo et al. (2022) conclude that the lack of social interaction while shopping online will discourage older consumers from doing so. Regarding spending habits, Truong and Truong (2022) find that with the increase in a consumer's age, the odds of spending more while shopping online increase, while the same decrease while shopping in-store.

The work status of an individual may as well determine their shopping habits. Due to its simplicity and time-saving advantages, online purchasing may become more common, particularly in households with full-time employment and young children (Frank and Peschel, 2020). According to Garín-Muñoz et al. (2022), most Spanish online consumers are employed. Furthermore, studies by López Soler et al. (2021) and Smith et al. (2022) supported these findings by demonstrating that full-time employees in European countries are likely to make online purchases.

Regarding attained *education level*, it is argued that people with a higher education level typically have more confidence when implementing new technologies. Therefore, educated individuals prefer online shopping to traditional in-store shopping (Van Droogenbroeck and Van Hove, 2017). In addition, Dominici et al. (2021) found that Belgian consumers who purchase groceries online are mainly highly educated. Truong and Truong (2022) found that higher educated consumers will tend to make online purchases but will be more considerate and spend less than consumers with lower levels of education.

Yahya and Sugiyanto (2020) showed that the *urbanisation level of the residence* is an important factor influencing the decision of whether to shop online or in-store.

Their results indicated that individuals who live in urban areas have a stronger tendency to shop online. These findings are consistent with those of AbdulHussein et al. (2022), whose survey used data from Canadian consumers, and those of Anderson and Srinivasan (2003). Residents of rural areas have low shopping accessibility and therefore choose to shop online for specific groups of products (Yousefi et al., 2023). This is in line with Hood et al. (2020), who state higher, but constant levels of online shopping from consumers in urban areas, whereas consumers from rural areas shop less frequently online, but if they do so, their spending is much higher.

Personal monthly income has also been shown to be one of the determining factors in the decision to shop online. Both Hood et al. (2020) and Giannakopoulou et al. (2022) have shown that higher-income individuals are more likely to shop online than in-store. In addition, a study of U.S. consumers by Duffy et al. (2022) confirmed the relationship between higher income and online purchases. Truong and Truong (2022) conclude that higher-income individuals will spend more while shopping in-store and less while shopping online than lower-income individuals.

Besides sociodemographic characteristics, *shopping mode-specific determinants* influence consumers' shopping decisions. In his influential work on the two-factor motivation theory, Herzberg examined 14 factors for job satisfaction and then divided them into two groups: motivators and hygiene factors (Herzberg et al., 2017). According to his findings, satisfaction, and dissatisfaction are on two separate continuums, i.e., independent of each other. Unlike motivators, hygiene factors do not lead to higher motivation, but their absence leads to dissatisfaction. Inspired by the intuitive correctness of his approach, we opted for a similar division of factors in our analysis of factors influencing consumers' attitudes and decisions when shopping online. Thus, the first group of factors, referred to as online shopping destimulators (OSD), contains variables that might discourage consumers from shopping online, making them similar to Herzberg's hygiene factors. The other group, online shopping stimulators (OSS), contains factors that could stimulate consumers to shop online. These two composite indicators were the first nodes of the decision-tree model in this paper. Online shopping characteristics such as convenience, privacy, promotion, and pricing, as well as delivery attributes such as the importance and influence of delivery time, fees, and reception, were part of the research paper of Dias et al. (2021), supporting the argument for choosing the OSS and OSD as a basis for shopping mode decision. Aw et al. (2021) determined a list of shopping mode-related factors influencing consumers' online shopping decisions: online search convenience, perceived usefulness of online reviews, immediate possession of products, and smart shopping perception while researching products online. Aw et al. (2021) and Hermes et al. (2022) point out the influence of perceived risks and trust in online shopping to also shape the shopping mode choice. Influencing factors for in-store shopping were the need for interaction and perceived helpfulness of in-store salespeople, the need for a touch of products, price comparison orientation, and product knowledge. Rathee and Rajain (2019) and Hermes et al. (2022) point out that the consumer's need for touch is also essential, posing a challenge to choose online over the in-store shopping mode. As per Rummo et al. (2022), the lack of social interaction discouraged online shopping. Loyalty to a store and habit-influenced behaviour will also affect consumers' decisions in shopping mode (Audrain-Pontevia and Vanhuele, 2016).

Analysing consumers' shopping behaviour proved to be challenging due to the (in) ability to collect relevant data and because of appropriate data analysis methods selection. The procedure of exploring and analysing vast amounts of data with the goal of identifying relevant patterns and trends is referred to as *data mining* (Song et al., 2001). *Decision trees* have become one of this field's most widely used and powerful techniques. The nodes and leaves of a decision tree produce an understandable hierarchical structure. A class is represented by each leaf of the tree, and an attribute is tested by each node. A categorisation applicable to all instances that reach that node is formed by organising cases in a decision tree from the root to that leaf node. There are numerous methods to build decision trees. One of the key differences between them is the ability to identify the feature that results in the best split of the source dataset. Depending on the algorithm used to create the decision tree, a splitting criterion can be established using a variety of metrics. The dataset is split until the previously defined stopping condition is met, or there are no more attributes (Quinlan, 1993).

The best-known method for constructing a decision tree is the *C&RT* (*Classification & Regression Tree*) *algorithm* (Breimann, 1984). However, subsequent publications (for example, Mingers, 1989) have shown that overfitting and selection bias are common problems when using this algorithm. For this reason, Strasser and Webber (1999) proposed a new set of algorithms, *Conditional Inference Trees* (*CTree*), based on a framework that combines recursive binary partitioning and permutation tests. Nevertheless, Gomes et al. (2020) compared the predictive power of these two algorithms, and their results suggest that the C&RT algorithm generally gives better results than the CTree algorithm for large datasets, while Bertsimas et al. (2022) concluded that the CTree algorithm gives the most stable results for smaller datasets.

Principal component analysis (PCA) (Bro and Smilde, 2014; Vidal and Sastry, 2016; Kherif and Latypova, 2020) is a practical algorithm that linearly transforms and scales complex data. During the PCA process, the number of components decreases while maximising the variance of explained data. Moreover, using PCA enables the isolation and quantification of patterns which can reveal connections between variables and samples or create the ability to form new hypotheses. The power of the PCA process to reduce the number of components has proven helpful before using decision-tree algorithms (Howley et al., 2006; Hu et al., 2009). Reduction of factors (independent variables) results in efficiency gains in data processing, decision-tree structure optimisation, and higher forecast accuracy of decision trees (Nasution et al., 2018), making PCA a very useful tool to be used before creating a decision tree.

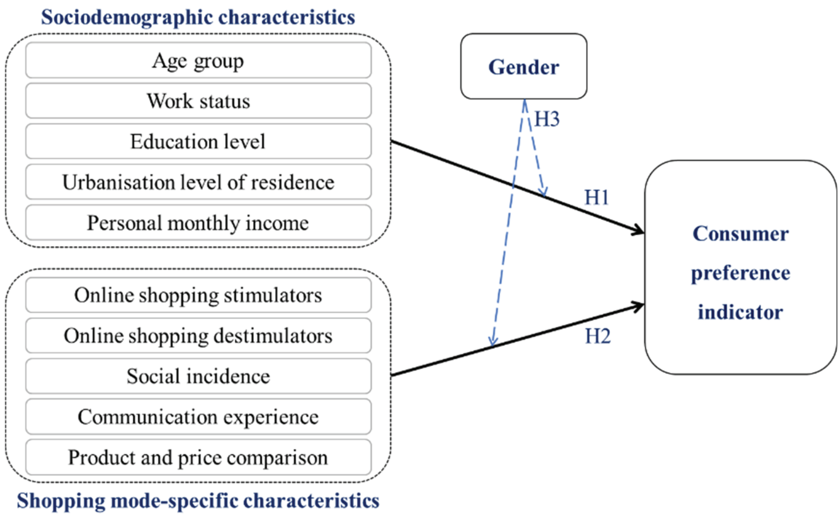
Recent research shows that decision trees are a valuable analytical tool for understanding consumers' behaviour. The C&RT and CTree algorithms have been widely used in consumer behaviour research. For example, Wah et al. (2011) used the C&RT algorithm to construct a decision tree to study consumers' car purchase behaviour and achieved 89.00% accuracy. On the other hand, Šebalj et al. (2017) studied in-store shopping intentions using several current decision-tree algorithms: J48, RandomForest, and REPTree (Reduced Error Pruning Tree). The results obtained from the study indicated that the *J48 algorithm* achieved the highest accuracy, with a classification rate of 84.75%. In addition, Ansari and Riasi (2019) also used the J48 tree algorithm to analyse consumers' preferences regarding shopping locations. The results of their study again showed that the J48 algorithm could be used as a strong tool for predicting consumers' behaviour due to its high accuracy. Furthermore, some authors studied shopping habits using PCA and decision trees. Significant results were presented in the work of Bucko et al. (2018), in which they used PCA to determine factors that influence consumers' purchase behaviour. Finally, Romano et al. (2014) successfully used PCA before constructing the decision tree to evaluate consumers' purchase decisions.

In assessing shopping behaviour with decision-tree algorithms using the mean values of multiple observed variables, some authors (Šebalj et al., 2017; Ansari and Riasi, 2019) constructed an indicator of shopping intention and shopping location preference indicator as binary variables. In their study of shopping mode preferences, Farag et al. (2007) found a relationship between the desired product type and the chosen shopping mode. In addition, Schmid and Axhausen (2018) examined in-store and online shopping preferences using the product groups of groceries and standard electronic devices, and Rossolov et al. (2021) did the same for a wide range of experiences and search goods. The collective results show that the shopping mode choice differs according to the product purchased. This suggests that the *consumer's preference indicator* can be built on evaluating the preferred shopping mode based on different product groups.

Based on the set working hypotheses and grounded in the conducted and presented literature review, we were able to shape a research model presented in Figure 1 and specify the following research hypotheses:

> *H1: Sociodemographic characteristics of consumers have a significant influence on a preferred shopping mode choice.*

Figure 1: Consumer's preference indicator influences



Source: Author's construction

While a specific age range of consumers portrays the highest tendency to shop online (Hood et al., 2020; Moon et al., 2021), it is generally the case that with an increase in age, consumers are more likely to shop in-store (Dias et al., 2022; Buhaljoti et al., 2022; Giannakopoulou et al., 2022). Higher educated consumers (Van Droogenbroeck and Van Hove, 2017; Dominici et al., 2021; Truong and Truong, 2022) and those with higher PMI (Hood et al., 2020; Giannakopoulou et al., 2022; Duffy et al., 2022) are more likely to shop online. Consumers' work status and time management (Frank and Peschel, 2020; Garín-Muñoz et al., 2021; López Soler et al., 2021; Smith et al., 2022), together with the urbanisation level of residence (Yahya and Sugiyanto, 2020; AbdulHussein et al., 2020; Yousefi et al., 2023) affect the consumer's shopping mode choice as well.

H2: Shopping mode-specific characteristics have a significant influence on a preferred shopping mode choice.

Importance of social interaction and staff helpfulness, physical product features, product comparison ability or shop loyalty (Audrain-Pontevia and Vanhuele, 2016; Rathee and Rajain, 2019; Hermes et al., 2022; Rummo et al., 2022), perceived risk and trust influence of online shopping (Aw et al., 2021; Hermes et al., 2022), online shopping convenience, information, review availability and privacy (Dias et al., 2022; Aw et al., 2021), all will influence the shopping mode choice.

H3: Male and female consumers have significantly different shopping mode decision patterns.

Men and women prefer to choose different shopping modes based on specific products purchasing (Kim et al., 2020; Yahya and Sugiyanto, 2020), with women's tendency to shop more online generally (Hood et al., 2020; Moon et al., 2021) and spend more while shopping online as well (Truong and Truong, 2022). Undesirable shopping mode aspects will also affect the genders' shopping mode choices differently (Dias et al., 2022).

3. Methodology

In order to gain better insight into influential factors while deciding on the purchase method, consumers answered a wide range of questions evaluating shopping specifics from different viewpoints. To make a large amount of gathered data more easily analysed, a principal component analysis was conducted with the aim of identifying comprehensive factors of influence (Bro and Smilde, 2014; Vidal and Sastry, 2016; Kherif and Latypova, 2020). Consumers evaluated their shopping preferences in buying different groups of products: *groceries, clothing and footwear, technical equipment,* and *gifts and presents*. Grounded in the papers of Šebalj et al. (2017) and Ansari and Riasi (2019), based on the respondents' evaluation results, a *consumer's preference indicator (CPI)* was constructed, which takes on the values 1 = consumer prefers shopping in stores, and 2 = consumer prefers shopping online. Since the dependent variable can only take the values of 1 or 2, the problem of determining the respondents' shopping preferences is presented as a classification problem. Three classification algorithms (C&RT, CTree, J48) were used to build the decision-tree models. The C&RT algorithm uses *Gini impurity* (Breimann, 1984) as a splitting criterion, which is defined as follows:

$$Gini\ impurity(D) = 1 - \sum_{i=1}^{n} {p_i}^2, \quad (1)$$

where D is a dataset that consists of n classes and p_i is the probability that an instance belongs to class i. The CTree algorithm, on the other hand, uses the permutation test framework to find the optimal binary split (Strasser and Webber, 1999). Moreover, the J48 algorithm, as an implementation of the C4.5 algorithm (Quinlan, 1996), uses *Gain Ratio* as a splitting criterion:

$$Gain\ Ratio(A) = \frac{Gain(A)}{Entropy(A)}, \quad (2)$$

where *Gain* and *Entropy* are defined as follows:

$$Gain(A) = Entropy(A) - \sum_{j=1}^{m} \frac{|D_j|}{|D|} * Entropy(D_j) \quad (3)$$

and

$$Entropy(A) = -\sum_{i=1}^{n} p_i log_2 p_i. \quad (4)$$

Here, D represents a given dataset, D_j represents the j-th subset of D and A represents a specific attribute. Moreover, p_i is the proportion of class i, which belongs to dataset D (Mitchell, 1997).

The algorithm that achieved the highest classification accuracy rate out of the three (C&RT, CTree, J48) is presented in the following section. To evaluate the performance of each of the three models constructed using C&RT, CTree, and J48 algorithms, the source dataset was first randomly divided into a training dataset and a test dataset in an 80:20 ratio, as this ratio is the most commonly used (Géron, 2019). The training dataset was subjected to a ten-fold cross-validation method – split into ten equally sized subgroups or folds, and the model was trained and validated ten times, each time using a different fold for validation and the remaining nine folds for training (Hastie et al., 2009; Han et al., 2012). Finally, the model's performance was tested on a separate, unseen test dataset. The previously described procedure allows for better generalisation by allowing independent evaluation of the model on data it has not encountered during training or validation (Nisbet et al., 2009). The analysis was performed using R software version 4.2.3, the *rpart*, *partykit*, and *RWeka* packages were utilised to construct the decision trees, respectively (Hornik et al., 2023; Hothorn et al., 2023; Therneau et al., 2022).

4. Empirical data and analysis

A Google Forms questionnaire was used for the empirical study to gather data on Croatian consumers' shopping behaviour. The questionnaire was active from May to September 2022, and 350 respondents gave valid answers, forming the sample for this research (Table 1). Most of the sample are women (72.1%), 80% of the respondents fall within the 21-50 age range, and more than half have acquired a bachelor's or higher education degree. Most respondents are employed or self-employed (entrepreneurs), with 75.2% of the sample stating having up to 1,200 EUR of personal monthly income (PMI). Additionally, 64.1% of the respondents reside in larger cities.

Table 1: Sample characteristics

Sociodemographic characteristic (variable operationalisation)	Share (in %)
Gender (GEN)	
Male (*0*)	27.9
Female (*1*)	72.1
Age group (AG)	
<21 years (*1*)	8.9
21-30 years (*2*)	28.2
31-40 years (*3*)	24.4
41-50 years (*4*)	27.3
51-60 years (*5*)	6.9
>60 years (*6*)	4.3
Work status (WS)	
Unemployed (*1*)	17.5
Employed pupil/student (*2*)	17.5
Entrepreneur (*3*)	10.9
Employed (*4*)	50.3
Retired (*5*)	3.7
Education level (EL)	
Lower education (*1*)	3.4
High school education (*2*)	44.5
Bachelor education (*3*)	15.8
Master or higher education (*4*)	36.2
Urbanisation level of residence (URB)	
Municipality with <10,001 inhabitants (*1*)	19.5
Municipality/city with 10,001-20,000 inhabitants (*2*)	16.4
City with 20,001+ inhabitants (*3*)	64.1
Personal monthly income (PMI)	
0-400 EUR (*1*)	26.4
401-800 EUR (*2*)	24.1
801-1,200 EUR (*3*)	24.7
1,201-1,600 EUR (*4*)	13.2
1,601-2,000 EUR (*5*)	3.7
2,001+ EUR (*6*)	7.8

Source: Author's calculation

Consumers evaluated shopping mode-specific (in-store and online) determinants on a scalar valuation basis (binary; three- or five-point scale). A principal component analysis was conducted (Romano et al., 2014; Bucko et al., 2018) on the mode-specific determinants to find seven factors (Kaiser-Meyer-Olkin test KMO = 0.753, Bartlett's test of sphericity $p<0.001$) cumulatively explaining 63.13% of the total variability of the data. The seven determined factors of influence are online shopping stimulators (OSS), online shopping destimulators (OSD), impulsive shopping indicators (ISI), price- and product-related indicators (PPI), social incidence importance indicators (SIII), communication experience importance indicators (CEII), and habits- and time-related indicators (HTI). An overview of the factors, the share of variance explained by each factor, and their high loading variables, together with loading and mean values and their range, are presented in Table 2.

Table 2: PCA result factors overview

Factor	% of variance	Loading variables	Loading	Mean	Min.	Max.
Online shopping destimulators (OSD)	19.535	Possible fraud	0.813	2.443	1	5
		Not fulfilling expectations	0.807	2.402	1	5
		Long delivery time	0.781	2.379	1	5
		Lack of digital literacy	0.748	1.566	1	5
		Products different from advertised	0.720	2.448	1	5
Online shopping stimulators (OSS)	10.683	Available information	0.761	3.612	1	5
		Safety	0.654	3.563	1	5
		Easy access to reviews	0.639	3.287	1	5
		Better deals	0.639	3.615	1	5
		Saving time	0.637	3.945	1	5
Impulsive shopping indicators (ISI)	8.186	Buying unnecessary products	0.860	2.718	1	5
		Impulsiveness	0.860	2.592	1	5
Price- and product-related indicators (PPI)	7.398	Perceived price awareness	0.804	4.440	1	5
		Chasing high discounts	0.737	4.011	1	5
		Product research	0.543	3.572	1	5
Social incidence importance indicators (SIII)	6.132	Behaviour when encountering crowds	0.781	1.718	1	3
		Behaviour while shopping	0.745	1.305	1	2
Communication experience importance indicators (CEII)	6.036	Staff kindness	0.773	4.529	1	5
		Communication with people	0.699	2.471	1	5
Habits- and time-related indicators (HTI)	5.210	Shopping in stores because of routine or urgency	0.852	3.195	1	5

Source: Author’s calculation

The constructed consumer's preference indicator was used as the dependent variable, while sociodemographic characteristics and mode-specific influencing factors served as independent variables for decision-tree construction. The dependent variable CPI was expressed as a binary variable with two classes (1 and 2). Of the 350 respondents, 64.37% preferred in-store shopping (denoted as 1), while the rest (35.63%) preferred online shopping (denoted as 2). The results of each algorithm are shown in Table 3, providing information on the size of the tree, the number of leaves, and the values of correctly and incorrectly classified instances.

Table 3: Classification algorithm results

Algorithm used	Number of leaves	Size of the tree	Correctly Classified Instances	Incorrectly Classified Instances
C&RT	13	25	63.768%	36.232%
CTree	5	9	65.217%	34.783%
J48	38	57	91.304%	8.696%

Source: Author's calculation

From the results, it can be observed that the J48 algorithm achieved the highest *classification accuracy rate* on the test dataset with 91.30% correctly classified instances. Detailed classification measures obtained on the test dataset for the decision tree created with the J48 algorithm are presented in Table 4. *Sensitivity* (*recall* or TPR – *true positive rate*) represents the ability of the model to positively classify a data record, while *specificity* (TNR – *true negative rate*) represents the ability to negatively classify a data record. *Precision* represents the proportion of positive predictions that are correct. To combine sensitivity and precision into one measure, we calculated the *F-score* (*F-measure*), a harmonic mean of sensitivity and precision. A high F-score (closer to 1) indicates the high accuracy of the model (Han et al., 2012). In addition, FPR and FNR represent *false positive rates* and *false negative rates*, respectively.

Table 4: Detailed classification measures calculated on the test dataset

Sensitivity	Specificity	Precision	FPR	FNR	F-score
0.911	0.917	0.953	0.083	0.089	0.932

Source: Author's calculation

To further illustrate the performance of the model, a *table of confusion* or *confusion matrix* for the test dataset is also presented (Table 5). The confusion matrix is a two-dimensional matrix with one dimension indexed by an object's true class and the other by the class supplied by the classifier. Larger values in the main diagonal and smaller values outside the diagonal represent good results (Han et al., 2012). Out of 45 respondents in the test dataset who prefer shopping in stores, a decision tree created using the J48 algorithm was able to classify 41 of them correctly. On the other hand, out of 24 respondents in the test dataset who prefer online shopping, 22 were correctly classified by the constructed decision tree.

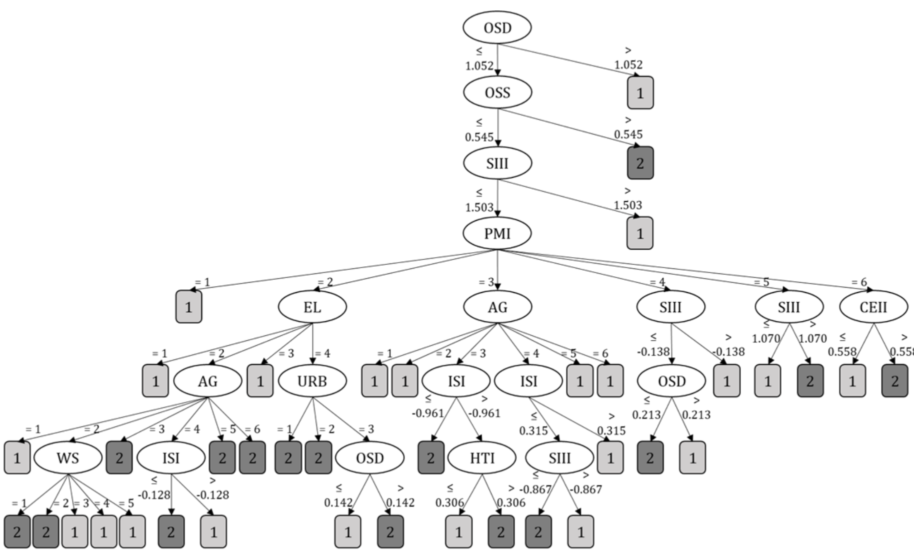
Table 5: Confusion matrix for the test dataset

Observed class	Predicted class	
	1	2
1	41	4
2	2	22

Source: Author's calculation

The structure of the decision tree, constructed using the J48 algorithm, is presented in Figure 2.

Figure 2: General consumers' preferences decision tree



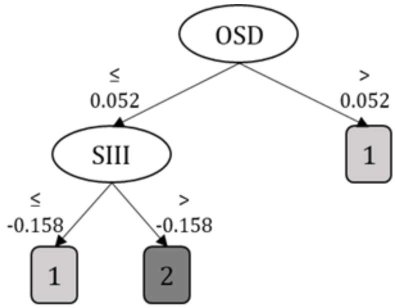
Source: Author's construction

We observe that the first major influencing factor to the preference is online shopping destimulators, where consumers strictly choose in-store shopping if the perceived level is high. Otherwise, consumers make decisions impacted by online shopping stimulators. If this indicator was moderately high, consumers preferred to shop online, but if lower, the importance of social incidence was considered. Consumers who commonly perform their shopping activity relaxed will choose to do so in shopping stores, even crowded ones. Even more so, some consumers based their decision on the attendance of shops, with more overcrowded shops seen as preferable.

Sociodemographic factors such as personal monthly income, education level, and age group also showed importance. Consumers from the lowest income group (0-400 EUR of personal monthly income – PMI) shopped in stores. The succeeding group with 401-800 EUR PMI decided differently based on their education level, although most preferred to shop online. Only lower-educated consumers employed consumers with a high school education, or consumers aged 40 and above with a higher level of impulsive shopping indicators were shopping in stores. Higher educated consumers within the second PMI level group preferred to shop online unless they lived in cities with over 20,000 inhabitants and had a moderate level of perceived online shopping destimulators. Age-related preference differences were present amongst consumers with an 801-1,200 EUR PMI level. Both the group's youngest consumers (30 years or less) and the oldest (51 years or more) chose to shop in stores. Consumers aged 31-50 were affected by their impulsive shopping behaviour, buying impulsive or unnecessary products online more often. If the social incidence importance indicator was higher, consumers aged 41-50 chose in-store purchases rather than online. Consumers with 1,201-1,600 EUR and 1,601-2,000 EUR PMI levels were further heavily affected by the social incident importance indicator but with different end-result levels. A higher social incidence indicator led the 1,201-1,600 EUR PMI level consumers to in-store shopping, while the 1,601-2,000 EUR PMI level consumers to online shopping. Lastly, consumers with over 2,000 EUR PMI within the group based their decision on the communication experience importance indicator, where somewhat unexpectedly, a higher level of staff kindness importance and communication with people was indicative of preferring to shop online.

In order to further analyse consumers' behaviour, we divided the original dataset into two smaller groups by gender and created decision trees using the J48 algorithm (as this again yielded the highest classification accuracy rate), separately for each of the subgroups. The decision tree in Figure 3 shows how male consumers choose their shopping mode, while the decision tree in Figure 4 shows what factors influence female consumers' online or in-store shopping. The detailed classification measures and confusion matrices obtained on the test dataset for these decision trees can be seen in Table 6. As shown in this table, the decision tree based on male data in Figure 3 correctly classified 85.00% of the instances, while the decision tree based on female data in Figure 4 correctly classified 94.12% of the instances.

Figure 3: Male consumers' preferences decision tree



Source: Author's construction

As can be seen by the decision tree (Figure 3), men consider much fewer factors when deciding to shop online or in-store. The first factor of influence on men's consumer preferences was online shopping destimulators, where a higher level proved to steer men towards in-store purchases. If the online shopping destimulators' influence level is not too high, men will decide based on the social incidence importance indicator. Men will choose in-store shopping unless the indicator is too negative, meaning the physical stores are too crowded, and men tend to want to leave the shop hurriedly, turning men to online shopping.

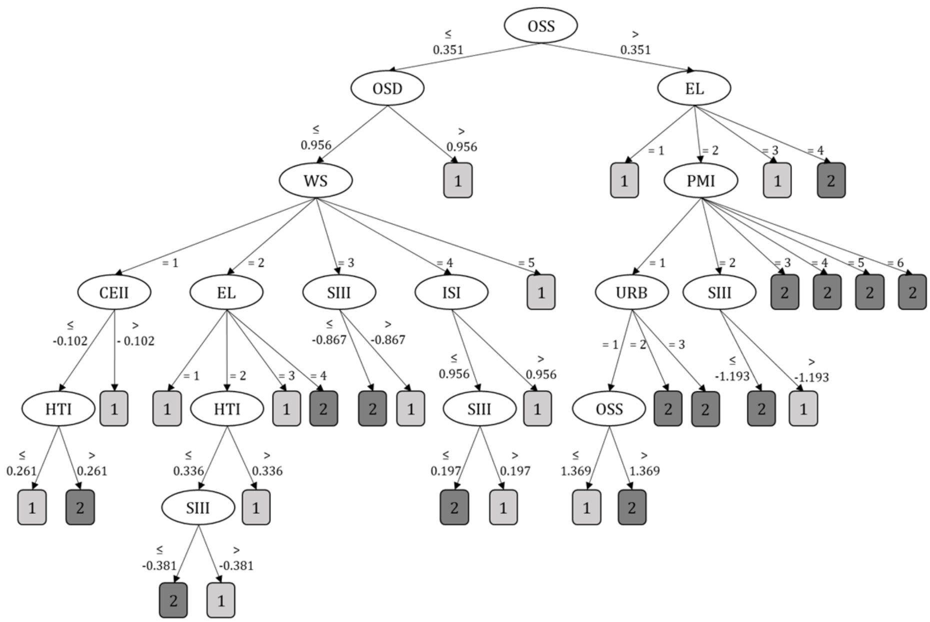
Table 6: Detailed classification measures and confusion matrices for the test dataset, male and female consumers' preferences decision tree

	Consumers		Confusion matrix		
	Male	Female	Predicted class	Observed class	
Number of leaves	3	29	*Male*	1	2
Size of the tree	5	45	1	13	0
Correctly Classified Instances	85.000%	94.118%	2	3	4
Incorrectly Classified Instances	15.000%	5.882%			
Sensitivity	1.000	0.969			
Precision	0.813	0.941			
Specificity	0.571	0.889	Predicted class	Observed class	
FPR	0.429	0.111	*Female*	1	2
FNR	0.000	0.030	1	32	1
F-score	0.897	0.955	2	2	16

Source: Author's calculation

Unlike men, women's consumer preferences are grounded firstly on online shopping stimulators (Figure 4). There is no level of influence of online shopping stimulators where women were categorically choosing online shopping. Instead, if the level rises to a moderate influence, the sociodemographic factors influence the final decision, such as education level, personal monthly income and urbanisation level of the residence. Lower educated women choose in-store shopping, and the most educated prefer online shopping. Women with a high school education decided differently based on their personal monthly income level. With up to 400 EUR of PMI, only the women living within smaller municipalities would shop in-store. Women with 401-800 EUR PMI level will choose more crowded stores, and women with 801 EUR and more PMI will prefer online shopping.

Figure 4: Female consumers' preferences decision tree



Source: Author's construction

If women's online shopping stimulator levels were lower, online shopping destimulators would determine their shopping preferences. If the level is moderately high, they will choose in-store shopping, but if lower than that, a wide range of factors affect their final decision. Sociodemographic factors like their work status and education level, or other mode-specific factors like the communication experience importance indicator, social incident importance indicator, impulsive shopping indicator and habits and time-related indicator proved to be of influence. The retired women preferred to shop in-store. The unemployed ones would make their decision based on the communication experience importance indicator. If communication and staff kindness were important, unemployed women would prefer to shop in-store. Still, if not, they would choose in-store shopping solely because of habit or time restrictions. Female employed pupils or students were influenced by attained education level. Once again, the lower educated preferred in-store, while the most educated selected online shopping. High school graduates, if time-restricted and under the heavy habitual influence, were to choose in-store purchases but would swap to online shopping if the stores were too crowded. Other female employees were influenced by impulsive shopping indicators, buying more impulsively in stores and choosing more crowded stores. Based on the social incidence importance indicator, female entrepreneurs decided to shop once again in better-visited stores.

5. Results and discussion

One of the most important takeaways from our study is a better understanding of the two main groups of factors that act in different directions, similar to the motivators and hygiene factors presented by Herzberg et al. (2017). Online shopping stimulators (OSS) are factors that drive consumers to shop online, while online shopping destimulators (OSD) work in the opposite direction, i.e., when destimulators are strong, consumers tend to make more traditional in-store purchases. Understanding these factors is even more important because they have been found to influence the first nodes of decisions in our tree model. The most important differences in consumers' decision-making are found between the genders, aligning with the research of Kim et al. (2020), Yahya and Sugiyanto (2020), Hood et al. (2020), Moon et al. (2021) and Trouong and Trouong (2022). Men show simpler, straightforward behaviour patterns, as they are found to shop in stores when they are not convinced of the safety of various aspects of online shopping, i.e., when their OSD is high. Moreover, they also shop in physical stores when their confidence in online shopping is high unless they are crowded. The influences of a larger number of factors shape women's decisions. High levels of online shopping stimulators (OSS) will entice them to shop online. In contrast, high levels of online shopping destimulators (OSD) will make them more likely to shop in physical stores, similar to men. Other factors influencing women's decisions include their education level, work status, income level, and the importance of social incidence. A higher tendency to shop online was found among a subgroup of women with the highest level of education. In addition, female entrepreneurs showed a higher tendency to shop online. At the same time, unemployed and retired women had a strong preference for shopping in stores, coming in line with the findings of Frank and Peschel (2020), Garín-Muñoz et al. (2021), López Soler et al. (2021) and Smith et al. (2022), who find that employed consumers generally are more likely to make online purchases. Similarly, higher-income women were found to be more likely to shop online. The social incidence importance indicator

(SII) showed that women prefer to shop in stores when they are crowded. It can be argued that women with higher levels of education, women in the business world, and women with higher incomes are more familiar with technology, including online shopping, and are therefore more likely to shop online. The same demographic influences should also be important for men, though our research showed that they base their decision on fewer factors.

In addition to gender differences, there are other general findings from this research. One of these was that educated consumers are more likely to shop online, building on the findings from Van Droogenbroeck and Van Hove (2017), Dominici et al. (2021) and Troung and Troung (2022). All income groups were likely to shop in stores when comparing different income levels. However, as Hood et al. (2020) and Giannakopoulou et al. (2022) conclude, online shopping becomes more likely as income increases. Similarly, across the different age groups, all age groups tend to buy in-store, but this tendency is more pronounced among the oldest and youngest consumers. The results show that middle-aged groups are the most likely to shop online, as in the papers from Hood et al. (2020) and Moon et al. (2021). Also, as might be expected, consumers with an impulsive urge to shop, as well as consumers who place a high value on social incidence, are most likely to shop in stores. One factor analysed in this study that proved influential on decision-making in prior research (Yahya and Sugiyanto, 2020, AbdulHussein et al., 2020, Yousefi et al., 2023) was the residence's urbanisation degree. In our model, this factor moderately influences the decision-making of respondents living in areas with different population densities but only combined with purchasing specific product types.

6. Conclusion

Is the glass half full or half empty? can be used as an analogy to this paper's results. Are consumers more influenced by online shopping stimulators or online shopping destimulators? Based on the results of the J48 algorithm, which gave the highest classification accuracy rate, generally, online shopping destimulators are a more decisive influence on consumers' preferences, supporting a set hypothesis H2 on shopping mode specifics influencing shopping decisions. Interestingly, the first significant difference between female and male consumers can be observed immediately. When deciding on their preferred shopping mode, men first consider online shopping destimulators, while women act based on their online shopping stimulators' views. The second significant difference between female and male consumers' preferences is the number of influencing factors affecting the final choice, confirming the set hypothesis H3 on gender differences in shopping behaviour. Men consider substantially fewer factors than women while deciding on the shopping mode. It can also be concluded that mode-specific factors are first in the line of influence on consumers' preferences. The sociodemographic characteristics of consumers also have a significant impact, verifying hypothesis H1 that sociodemographic subgroups act differently while shopping. Still, they appear secondary, most notable being consumers' work status, education level, and age, but other mode-specific shopping factors further influence different sociodemographic subgroups' consumer preferences. The presented new knowledge is beneficial to managers in the trading sector, as they can adapt their business to shifts in consumer behaviour post a weighty market disruption period, maximising the market potential of consumer niches. We need to emphasise that the research sample size and proportion of men and women partaking may limit the general inferences. Apart from the relative size of our sample, which limits the strength and scope of our findings, it would be beneficial for future research to include respondents from neighbouring countries with a higher percentage of consumers with the online shopping experience. This would allow us to determine whether differences in attitudes toward online shopping are due to respondents' demographics or geography. In addition, one of the limitations of this study is due to the questionnaire, as respondents who had no prior experience with online shopping were not asked to answer questions about their attitudes toward online shopping. If this were not the case, a better understanding of their dislike could be achieved. This limitation should also be considered in our future efforts to understand shopper decision-making better.

This paper is supported through project ZIP-UNIRI-2023-14 by the University of Rijeka

References

AbdulHussein, A., Cozzarin, B., Dimitrov, S. (2022) "Changes in consumer spending behavior during the COVID-19 pandemic across product categories", *Electron Commerce Research*, https://doi.org/10.1007/s10660-022-09618-9.

Anderson, R. E., Srinivasan, S. S. (2003) "E-satisfaction and e-loyalty: A Contingency Framework", *Psychology & Marketing*, Vol. 20, No. 2, pp. 123–138, https://doi.org/10.1002/mar.10063.

Ansari, A., Riasi, A. (2019) "Using Decision Trees to Analyse the Customers' Shopping Location Preferences", *International Journal of Business Excellence*, Vol. 18, No. 2, pp. 174–202, https://doi.org/10.1504/ijbex.2019.099557.

Audrain-Pontevia, A.-F., Vanhuele, M. (2016) "Where do Customer Loyalties Really Lie, and Why? Gender Differences in Store Loyalty", *International Journal of Retail & Distribution Management*, Vol. 4, No. 8, https://doi.org/10.1108/ijrdm-01-2016-0002.

Aw, E. C.-X. et al. (2021) "Unravelling Determinants of Webrooming Behavior: A Qualitative Inquiry", *International Journal of Business and Society*, Vol. 22, No. 3, https://doi.org/10.33736/ijbs.4321.2021.

Bertsimas, D., Dunn, J., Gibson, E. (2022) "Optimal Survival Trees", *Machine Learning* 111, pp. 2951–3023, https://doi.org/10.1007/s10994-021-06117-0.

Breimann, L. (1984) *Classification and Regression Trees,* 1st Edition, New York: Routledge.

Bro, R., Smilde, A. K. (2014) "Principal Component Analysis", *Analytical Methods*, Vol. 6, No. 9, pp. 2812–2831, https://doi.org/10.1039/c3ay41907j.

Bucko, J., Kakalejčík, L., Ferencová, M. (2018) "Online Shopping: Factors That Affect Consumer Purchasing Behaviour", *Cogent Business & Management*, Vol. 5, No. 1, https://doi.org/10.1080/23311975.2018.1535751.

Buhaljoti, A., Habili, M., Abazi, A. (2022) "The Impact of Knowing the Profile of Online Shoppers on Online Shopping: Evidence from City of Berat, Albania", *WSEAS Transactions on Business and Economics*, Vol. 19, https://doi.org/10.37394/23207.2022.19.112.

Dias, E. G., Oliveira, L. K., Isler, C. A. (2022) "Assessing the Effects of Delivery Attributes on E-Shopping Consumer Behaviour", *Sustainability*, Vol. 14, No. 1, https://doi.org/10.3390/su14010013.

Dominici, A. et al. (2021) "Determinants of Online Food Purchasing: The Impact of Socio-demographic and Situational Factors", *Journal of Retailing and Consumer Services*, Vol. 60, https://doi.org/10.1016/j.jretconser.2021.102473.

Duffy, E. W. et al. (2022) "Prevalence and Demographic Correlates of Online Grocery Shopping: Results from a Nationally Representative Survey during the COVID-19 Pandemic", *Public Health Nutrition*, Vol. 25, No. 11, pp. 3079–3085, https://doi.org/10.1017/S1368980022001756.

Eurostat (2023-a) *Individuals having ordered/bought goods or services for private use over the internet in the last three months.* Available at: <https://ec.europa.eu/eurostat/databrowser/view/TIN00067/default/table?lang=en&category=isoc.isoc_i.isoc_iec> [Accessed: March 10, 2023].

Eurostat (2023-b) *Internet purchases – perceived barriers (2021 onwards).* Available at: https://ec.europa.eu/eurostat/databrowser/view/ISOC_EC_INB21/default/table?lang=en&category=isoc.isoc_i.isoc_iec [Accessed March 10, 2023].

Eurostat (2023-c) *Internet purchases – problems encountered (2021 onwards).* Available at: <https://ec.europa.eu/eurostat/databrowser/view/ISOC_EC_IPRB21/default/table?lang=en&category=isoc.isoc_i.isoc_iec> [Accessed March 10, 2023].

Farag, S. et al. (2007) "Shopping Online and/or in-store? A Structural Equation Model of the Relationships Between e-shopping and in-Store Shopping", *Transportation Research Part A: Policy and Practice*, Vol. 41, No. 2, pp. 125–141, https://doi.org/10.1016/j.tra.2006.02.003.

Frank, D.-A., Peschel, A. O. (2020) "Sweetening the Deal: The Ingredients that Drive Consumer Adoption of Online Grocery Shopping", *Journal of Food Products Marketing*, Vol. 26, No. 8, pp. 535–544, https://doi.org/10.1080/10454446.2020.1829523.

Garín-Muñoz, T., Amaral, T., Valarezo-Unda, A. (2022) "Evolution of the Internet Gender Gaps in Spain and Effects of the Covid-19 Pandemic", *Telecommunications Policy*, Vol. 46, No. 8, https://doi.org/10.1016/j.telpol.2022.102371.

Géron, A. (2019) *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, O'Reilly Media, Inc.

Giannakopoulou, M., Adamides, G. & Stylianou, A. (2022) "Consumer Adoption of Online Grocery Shopping: Findings from a Case Study in Cyprus". In *Proceedings of the 10th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2022)*, September 22-25, Athens, Greece, CEUR Workshop Proceedings 3293, pp. 164–171.

Gomes, C., Lemos, G. C., Jelihovschi, E. (2020) "Comparing the Predictive Power of the CART and CTREE Algorithms", *Avaliação Psicológica*, Vol. 19, No.1, pp. 87–96, https://doi.org/10.15689/ap.2020.1901.17737.10.

Han, J., Kamber, M., Pei, J. (2012) *Data Mining Concepts and Techniques*, 3rd Edition, Morgan Kaufmann.

Hastie, T., Tibshirani, R., Friedman, J. H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition, Springer.

Hermes, A. et al. (2022) "Exploring Online and In-Store Purchase Willingness: Associations with the Big Five Personality Traits, Trust, and Need for Touch", *Frontiers in Psychology*, Vol. 13, pp. 1–17, https://doi.org/10.3389/fpsyg.2022.808500.

Herzberg, F., Mausner, B., Block Snyderman, B. (2017) *The Motivation to Work*, New York: Routledge, Oxon.

Hood, N. et al. (2020) "Sociodemographic and Spatial Disaggregation of e-commerce Channel Use in the Grocery Market in Great Britain", *Journal of Retailing and Consumer Services*, Vol. 55, https://doi.org/10.1016/j.jretconser.2020.102076.

Hornik, K., et al. (2023) *Package RWeka*. Available at: <https://CRAN.R-project.org/package=RWeka> [Accessed: May 13, 2023].

Hothorn, T., Seibold, H., Zeileis, A. (2023) *Package partykit*. Available at: <https://CRAN.R-project.org/package=partykit> [Accessed: May 13, 2023].

Howley, T. et al. (2006) "The Effect of Principal Component Analysis on Machine Learning Accuracy with High-dimensional Spectral Data", *Knowledge-Based Systems*, Vol. 19, No. 5, pp. 363–370, https://doi.org/10.1016/j.knosys.2005.11.014.

Hu, J., Deng, J. & Sui, M. (2009) "A New Approach for Decision Tree Based on Principal Component Analysis". In *2009 International Conference on Computational Intelligence and Software Engineering*, 11-13 December, Wuhan, China, Institute of Electrical and Electronics Engineers, https://doi.org/10.1109/cise.2009.5366006.

Kherif, F., Latypova, A. (2020) "Principal component analysis", In Michelli, A. and Vieira, S. ed., *Machine Learning: Methods and Applications to Brain Disorders*, Academic Press, pp. 209–225, https://doi.org/10.1016/b978-0-12-815739-8.00012-2.

Kim, J.-H. et al. (2020) "Consumer Decision-making in a Retail Store: The Role of Mental Imagery and Gender Difference", *International Journal of Retail & Distribution Management*, Vol. 49, No. 3, https://doi.org/10.1108/ijrdm-10-2019-0353.

López Soler, J. R., Christidis, P., Vassallo, J. M. (2021) "Teleworking and Online Shopping: Socio-Economic Factors Affecting Their Impact on Transport Demand", *Sustainability*, Vol. 13, No. 13, https://doi.org/10.3390/su13137211.

Mingers, J. (1989) "An Empirical Comparison of Selection Measures for Decision-Tree Induction", *Machine Learning*, Vol. 3, pp. 319–342, https://doi.org/10.1023/A:1022645801436.

Mitchell, T. M. (1997) *Machine Learning*, McGraw Hill Science/Engineering/Math.

Moon, JH., Yunseon C., HakJun S. (2021) "Determinants of Consumers' Online/Offline Shopping Behaviours during the COVID-19 Pandemic", *International Journal of Environmental Research and Public Health*, Vol. 18, No. 4, https://doi.org/10.3390/ijerph18041593.

Nasution, M. Z. F., Sirompul, O. S., Ramli, M. (2018) "PCA Based Feature Reduction to Improve the Accuracy of Decision Tree c4.5 Classification", *Journal of Physics: Conference Series*, Vol. 978, No. 1, https://doi.org/10.1088/1742-6596/978/1/012058.

Nisbet, R., Elder, J., Miner, G. (2009) *Handbook of Statistical Analysis & Data Mining Applications*, Academic Press, https://doi.org/10.1016/B978-0-12-374765-5.X0001-0.

Our World in Data (2022) *Landline Internet subscriptions per 100 people,* 2020. Available at: <https://ourworldindata.org/grapher/broadband-penetration-by-country?time=latest&region=Europe> [Accessed: March 8, 2023].

Quinlan, J. R. (1993) *C4.5: Programs for Machine Learning*, Los Angeles, CA: Morgan Kaufmann.

Quinlan, J. R. (1996) "Improved Use of Continuous Attributes in C4.5.", *Journal of Artificial Intelligence Research*, Vol. 4, pp. 77–90, https://doi.org/10.1613/jair.279.

Rathee, R., Rajain, P. (2019) "Online shopping environments and consumer's Need for Touch", *Journal of Advances in Management Research*, Vol. 14, No. 5, https://doi.org/10.1108/jamr-12-2018-0116.

Romano, R., Davino, C., Næs, T. (2014) "Classification Trees in Consumer Studies for Combining both Product Attributes and Consumer Preferences with Additional Consumer Characteristics", *Food Quality and Preference*, Vol. 33, pp. 27–36, https://doi.org/10.1016/j.foodqual.2013.11.006.

Rossolov, A., Rossolova, H., Holguín-Veras, J. (2021) "Online and in-store Purchase Behaviour: Shopping Channel Choice in a Developing Economy", *Transportation*, Vol. 48, pp. 3143–3179, https://doi.org/10.1007/s11116-020-10163-3.

Rummo, P. E. et al. (2022) "Age-Specific Differences in Online Grocery Shopping Behaviours and Attitudes among Adults with Low Income in the United States in 2021". *Nutrients*, Vol. 14, No. 20, https://doi.org/10.3390/nu14204427.
Schmid, B., Axhausen, K. W. (2018) "In-store or Online Shopping of Search and Experience Goods: A Hybrid Choice Approach", *Journal of Choice Modelling*, Vol. 31, pp. 156–180, https://doi.org/10.1016/j.jocm.2018.03.001.

Smith, L. G. et al. (2022) "Comparing Household and Individual Measures of Access through a Food Environment Lens: What Household Food Opportunities Are Missed When Measuring Access to Food Retail at the Individual Level?", *Annals of the American Association of Geographers,* Vol. 112, No. 2, https://doi.org/10.1080/24694452.2021.1930513.

Song, H. S., Kim, J. K., Kim, S. H. (2001) "Mining the Change of Customer Behaviour in an Internet Shopping Mall", *Expert Systems with Applications*, Vol. 21, No. 3, pp. 157–168, https://doi.org/10.1016/S0957-4174(01)00037-9.

Strasser, H., Weber, C. (1999) *On the Asymptotic Theory of Permutation Statistics*, WU Working Paper, Report Series SFB "Adaptive Information Systems and Modelling in Economics and Management Science" No. 27, *SFB Adaptive Information Systems and Modelling in Economics and Management Science*, WU Vienna University of Economics and Business.

Šebalj, D., Franjković, J., Hodak, K. (2017) "Shopping Intention Prediction Using Decision Trees", *Millenium – Journal of Education, Technologies, and Health*, Vol. 2, No. 4, pp. 13–22. https://doi.org/10.29352/mill0204.01.00155.

Therneau, T., Atkinson, B., Ripley, B. (2022) *Package rpart*. Available at: <https://CRAN.R-project.org/package=rpart> [Accessed: May 13, 2023].

Truong, D., Truong, M. D. (2022) "How do Customers Change their Purchasing Behaviors During the COVID-19 Pandemic?", *Journal of Retailing and Consumer Services,* Vol. 67, https://doi.org/10.1016/j.jretconser.2022.102963.

Van Droogenbroeck, E., Van Hove, L. (2017) "Adoption of Online Grocery Shopping: Personal or Household Characteristics?", J*ournal of Internet Commerce*, Vol. 16, No. 3, https://doi.org/10.1080/15332861.2017.1317149.

Vidal, R., Ma, Y., Sastry, S. S. (2016) *Principal Component Analysis*. In Vidal, R., Ma, Y., Sastry, S. S. ed., *Generalized Principal Component Analysis*, Springer New York, NY, pp. 25–62, https://doi.org/10.1007/978-0-387-87811-9_2.

Yahya, S., Sugiyanto, C. (2020) "Indonesian Demand for Online Shopping: Revisited", *Journal of Indonesian Economy and Business*, Vol. 35, No. 3, pp. 188–203, https://doi.org/10.22146/jieb.55358.

Yousefi, N., Wang, K., Circella, G. (2023) "Factors Influencing the Types of Merchandise Purchased Online: Evidence from the 2018 California Survey of Emerging Transportation Trends", *Transportation Research Interdisciplinary Perspectives*, Vol. 17, https://doi.org/10.1016/j.trip.2022.100734.

Wah, Y. B., Ismail, N. H., Fong, S. (2011) "Predicting car purchase intent using data mining approach". In *Proceedings of Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 26-28 July, Shanghai, China, Institute of Electrical and Electronics Engineers, pp. 1994–1999.

Stabla odlučivanja ne lažu: Zanimljivosti o sklonostima hrvatskih online potrošača

Ana Marija Filipas[1], *Nenad Vretenar*[2], *Ivan Prudky*[3]

Sažetak

Razumijevanje preferencija potrošača uvijek je bilo važno kako za ekonomsku teoriju, tako i za operacijski menadžment, upravljanje lancem opskrbe, marketing i druge poslovne aktivnosti. Iako se preferencije u teorijskom modeliranju često smatraju stabilnima, to nije slučaj u donošenju odluka u stvarnom svijetu. Zbog toga je preferencije potrošača potrebno dobro razumjeti, a to posebno vrijedi u okolnostima kada dolazi do poremećaja na tržištu. Cilj ovog istraživanja bolje je razumijevanje preferencija potrošača u internet kupovini nakon što su tržišta bila pogođena pandemijom COVID-19. U tu svrhu, provedeno je anketno istraživanje među hrvatskim potrošačima s prethodnim iskustvom u internet kupovini. Upitnik je bio aktivan od svibnja do rujna 2022. godine, a ispunilo ga je 350 ispitanika. S ciljem ustanovljavanja utjecaja analiziranih čimbenika kupovine i demografskih karakteristika potrošača na ponašanje potrošača, korišteni su modeli stabla odlučivanja uz pomoć tri klasifikacijska algoritma. Od tri odabrana algoritma, algoritam J48 je na testnim podacima pokazao najveću stopu točnosti u klasifikaciji. Glavne komponente pokazatelja koji utječu na ponašanje potrošača u ovom istraživanju su stimulatori i destimulatori internet kupovine, a zatim i važnost društvene incidencije. Rezultati istraživanja upućuju na značajne razlike u potrošačkim navikama između muškaraca i žena, pri čemu muškarci koriste manje varijabli pri donošenju odluka. Osim toga, analiza donošenja odluka pokazala je razlike i prema nizu drugih demografskih čimbenika te razlike kod kupnje različitih grupa proizvoda.

Ključne riječi: *teorija odlučivanja, preferencije potrošača, rudarenje podataka, stabla odlučivanja, pokazatelji ponašanja potrošača*

JEL klasifikacija: *C44, D12*

[1] *Asistent, Sveučilište u Rijeci, Ekonomski fakultet, Ivana Filipovića 4, 51 000 Rijeka, Hrvatska. Znanstveni interes: kvantitativna ekonomija, ekonometrija, statistika. Tel.: +385 51 355 162. E-mail: ana.marija.filipas@efri.hr.*

[2] *Izvanredni profesor, Sveučilište u Rijeci, Ekonomski fakultet, Ivana Filipovića 4, 51 000 Rijeka, Hrvatska. Znanstveni interes: teorija organizacije, operacijski menadžment, teorija odlučivanja. Tel.: +385 51 355 163. E-mail: nenad.vretenar@efri.hr.*

[3] *Asistent, Sveučilište u Rijeci, Ekonomski fakultet, Ivana Filipovića 4, 51 000 Rijeka, Hrvatska. Znanstveni interes: kvantitativna ekonomija, teorija odlučivanja, ekonomika sporta. Tel.: +385 51 355 161. E-mail: ivan.prudky@efri.hr.*