Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2022 doi:10.20944/preprints202208.0148.v1

The effect of the number of e-stores subscribers on Chinese smartphone brand purchases: Evidence from a machine learning model.

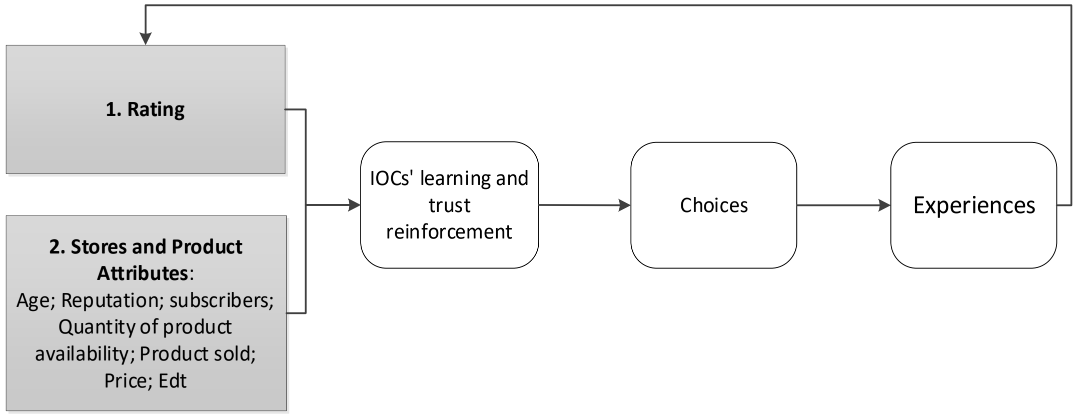
Karamoko K.E.H. N'da [1]*, Jiaoju Ge [2], Steven Ji-Fan Ren [3] and Jia Wang [4]
School of Economics and Management, Harbin Institute of Technology (Shenzhen), Shenzhen, China, 518055; andabuy@yahoo.com
School of Economics and Management, Harbin Institute of Technology (Shenzhen), Shenzhen, China, 518055; jiaoge@hit.edu.cn
School of Economics and Management, Harbin Institute of Technology (Shenzhen), Shenzhen, China, 518055; renjifan@hit.edu.cn
School of Business Administration, Smart Cities Research Institute, American University of Sharjah, United Arab Emirates, xiaobo@aus.edu
*Correspondence: andabuy@yahoo.com

Abstract
Introduction*. Until now, the impact of learning variables on consumers' choices concerning Chinese product brands in the international online shopping framework remains unknown. Accordingly, this study aims to examine the effect of those learning variables on global consumers' choices of Chinese product brands.*
Method. *A total of 44,704 transactions related to the buying process have been collected from a programming language and the Octopus Software within a Chinese International Online Shopping platform.*
Analysis. *The 44,704 transactions have been analyzed through a Decision Tree.*
Results. *The study points out that the number of e-retailers' subscribers reinforces the international consumers' trust online. At the same time, the pricing levels and quantity of product availability are used by global online consumers to assess the originality of Chinese product brands.*
Conclusions*. First, this study extends the existing literature on consumer learning by going beyond the learning variables considered. Second, the study boosts consumer learning literature by elucidating the most significant learning variables guiding international online consumers' choices and purchases. The application of the results will enable brands and e-retailers to understand (1) the stages of the international online consumers' choice; (2) the buying strategies of global consumers.*
Keywords: *Chinese smartphone brands, Decision trees, e-stores subscribers, consumer learning*

1. Introduction

Purchasing online is nowadays the norm in consumers' daily shopping modes because of the great convenience it offers (Hong *et al.*, 2017). In this perspective, Chinese domestic electronic commerce's rapid growth, supported by China's exceptional digital development, has opened up new prospects for its International online sales (IOS). The rate of Chinese global online sales in China's international trade increased from about 10 in 2010 to 40 in 2015 (Wang *et al.* 2017). In 2020 it was expected that the turnover would reach ¥ 12 trillion, representing 37.6% of Chinese international commerce (Wang *et al.*, 2017). This unprecedented growth relies on companies like Alibaba, which attract world consumers influenced by Chinese product brands' affordable prices. Arousing, thus, a growing interest among international online consumers in purchasing Chinese product brands directly via Chinese global online sales platforms. However, with technological development modifying consumer behaviour and purchase strategies (Constantinides, 2004), consumers' learning about transaction factors through information processing is increasingly considered (Arnould *et al*., 2004).
Indeed, the transaction process of IOS is quite different from that of domestic online shopping due to the geographical situation separating purchasers and sellers. Unlike national online shopping, customers and vendors live in various geographic areas and temporal. Besides, there is a high risk of purchasing fake products (Lin et al., 2018). Moreover, the refund process is more complicated than domestic online shopping (Lin et al., 2018). Therefore, international online consumers worry during the selection of the transactional components (e-retailers, product brands, and shipping conditions). To reduce risk-taking, buyers learn about the transaction components through buyers'

ratings and other variables related to the transaction, such as store age, store reputation, stores number of subscribers, Quantity of product availability, the number of products sold, price and delivery time (Edt). According to Mo et al. (2015), these buyers' ratings furnish post-purchase information, affecting new customers' purchasing behaviour and decision-making. Besides that, other attributes related to the transaction provide information on the state of the transaction and influence consumer behaviours. In this regard, Wang et al. (2017) state that product-related information, for instance, catch consumers' attention and impact their buying actions. Nevertheless, the effect of those learning variables on the choice and purchase of Chinese product brands remains unknown. Indeed, even though previous research has examined the impact of buyers' ratings on purchases and choices, very few studies have considered the combined effect of rating and transactional variables related to the products and stores (Zhao et al., 2013; Liu, 2006). Consequently, this study investigates the effect of these learning variables on the choice and purchase of Chinese smartphone brands.
Several works modelled the effect of some of those variables on purchase and choice under various purchase conditions using theories, such as the Bayesian learning model (e.g., Iyengar et al., 2007) and Stimulus-Organism-Response Model (e.g. Mo et al., 2015). However, this study goes beyond previous ones. In this study, we investigate the combined effect of buyers' ratings and variables related to products and stores on purchasing and choosing Chinese smartphone brands through a machine learning model.
The framework of this study concerns only the purchase and choice preferences of Chinese smartphone brands. When purchasing online products such as smartphones, their quality is unknown until consumers receive and use them. When purchasing online products such as smartphones, their quality is unknown until consumers receive and use them. It is not easy to be sure about the actual state of such a product before one has received it. Therefore, learning is highly recommended during the purchase process. Figure 1 presents the general framework of the study.
The study is organized in this manner: We first start with the literature review. Then we highlight the theoretical foundation of the previous work and present the decision tree algorithm's theoretical basis. Next, we explain the data structure and expose the results and discussions. Finally, we offer theoretical contributions, conclusions, implications, and limitations.



IOCs = international online consumers

Fig.1. General framework

2. Literature review

With a digital era reinforcing relationships between customers and products, consumers' learning before purchase is increasingly considered. Learning before purchase helps consumers cope with increased product complexity while affecting their attitudes, behaviours and preferences (Arnould et al., 2004; Fang and Xu, 2011). Many studies utilized various models and theories to provide insight into the effect of consumers learning on purchases. One may cite the Bayesian learning model, the most utilized model among those theories and models. Thus, using the bayesian model, Mehta et al. (2003) investigated the pricing search effect on the purchase when there is uncertainty on brand prices and found that consumers spend huge costs on brand prices search. However, the search costs do not affect the purchases. Erdem et al. (2004) discovered that learning about quality and price significantly affects brand choices in various countries. Iyengar et al. (2007) modelled consumers' learning of quality and quantity of services while taking on the base of the pricing scheme. They showed that consumers' learning of service plan quantity and quality based on the pricing scheme affects the service plan's choice. Zhu and Zhang (2010) have highlighted a differential impact of consumers' learning from consumer reviews in purchasing products in the same category. Zhao et al. (2011) modelled consumers' learning of product quality in the context of product degradation. Their results demonstrated that the context of product degradation affects consumers' purchase behaviour. Finally, Zhao et al. (2013) modelled the effect of consumer learning through online product reviews on consumer choice. The results show that favorable and frequent reviews have less impact on consumer choice.
Given the above, this research aims to study the effect of other learning factors, such as the ratings, the number of products available for sale, etc. These attributes provide information on the transaction's state. i.e., the product quality and e-store reputation can modify consumers' decision-making before purchasing. Table 1 below summarizes the above research.

Table.1. Research related to consumer learning and results

Year	Author	Data	Product	Choice	Learning Variable	Framework	Results
2003	Mehta et al.	Scanner data	Liquid detergents	Products	Price	Offline	Influences consumer purchase
2004	Erdem et al.	Scanner panel data	laundry detergent, toilet paper, margarine	Brands	Quality and price	Offline	Influences consumer purchase
2007	Iyengar et al.	Scanner panel data	Services	Plans of services	Quality, quantity and pricing scheme	Offline	Influences consumer choice
2010	Zhu and Zhang	Panel data	Video game	Products in the same category	Consumer reviews	Online	A differential impact on consumer purchase.
2011	Zhao et al.	Scanner panel data	Peanut butter in the context of a product degradation crisis	Products	Quality	Offline	The situation of product degradation affects consumers' purchase behavior.
2013	Zhao et al.	Scanner panel data	Book titles	Experiential products	Online product reviews	Online	Positive and frequent reviews have less impact on consumer choice

2.1. Learning variables and purchase decision

According to Lakshmanan and Krishnan (2011), consumer learning is essential in consumer decision-making. It provides consumers with information that influences their purchase attitudes and choice intentions. Arnould et al. (2004) consider it a series of actions the consumer takes during the purchasing process. Those actions can modify their point of view and purchase behaviour due to online information treatment. In line with that, we define in this study as learning variables any transactional variable susceptible to providing information on the transaction and able to impact consumers' purchases and choices. Therefore, we considered in this study two groups of learning variables. Namely, buyers' ratings and variables related to the transactional components such as products (e.g., price), e-retailers (e.g., subscribers), and shipping conditions (e.g., delivery time).

2.2. Consumer experiences: ratings

Technological development has modified consumer buying behaviour and strategies (Constantinides, 2004). Henceforth, online buyers can easily access former buyers' buying experiences (Zhao et al., 2013) and e-retailer selling behaviours. Research proves that these experiences impact purchases (e.g., Zhao et al., 2013; Chevalier and Mayzlin, 2006), consumers' conduct and choices (Kostyra et al., 2016). According to Mo et al. (2015), reviewing the ratings constitutes the initial step in the learning process. With these ratings, consumers can obtain former buyers' views concerning the transactional components. In tis regard, Mo et al. (2015) showed a substantial positive impact of ratings on online consumers' buying behaviours. Therefore, we incorporated them as learning variables to study their effect on IOCs' purchases and choices of Chinese smartphone brands.

2.3. Learning variables of the transaction components

Similar to domestic e-commerce, international online selling is a process that involves at least three phases on the selling platforms: the choice of the e-store outlet, product choice, and shipping conditions. Accordingly, each of these phases imposes decision-making on customers. Thus, once the customer decides to buy a particular product (Yrj and Saarij, 2019), he searches for the relevant information related to those 3 phases or components and then chooses the base of the available alternatives. The customer analyzes the attributes related to these three components before purchasing using the competitive advantage between different e-retailer outlets on the international platform. Therefore, adding the information about former buyers' ratings becomes essential for studying the effect of the learning variables on choice and purchase. Indeed, even though previous research has examined the impact of some learning variables, such as buyers' ratings (e.g., Zhao et al., 2013; Liu, 2006) about product purchases, few studies have considered, in addition to the buyers' ratings, the variables associated to products and stores.

2.4. Selling experience

There is a huge competition between different e-retailer outlets within the Chinese international online selling platforms. These selling platforms are a gateway between Chinese brands and global online consumers. In this context, the selling behavior of e-retailer outlets can influence consumers' attitudes and the reputation of Chinese brands. To deal with that, e-retailer outlets use their ratings to attract buyers. Yrj and Saarij (2019) show that retailer ratings are critically important in attracting customer purchases. Accordingly, we utilize the age of the e-retailer outlets as an indicator of their selling experience. Therefore, we included that variable as a learning variable.

2.5. Price

The price has always been considered a significant factor in product purchasing. In the international online shopping framework, consumers also pay customs duties in addition to the product price. However, as the customs fees and customs process differ from country to country, we did not consider the custom fee. Consequently, we considered only the product price. Some works have already investigated the effect of price learning on choices and purchases. For instance, Mehta et al. (2003), Iyengar et al. (2007), and Erdem et al. (2004) have all highlighted a positive influence of price learning on purchases. Therefore, the study included the price as a learning variable to study its effect on international online consumers' choice preferences between Chinese mobile phone brands.

2.6. Quantity of product available

The availability or not of a product at store outlets is a characteristic that one observes within numerous sale cases (Conlon et al., 2012). According to Conlon et al.(2012), that issue is among consumers' major worries in their purchase process. Most studies on product availability showed positive links between product availability and sales. However, the demand models have always been built by ignoring this fact (Shah et al., 2015). Similar to this, Shah et al. (2015) showed that any purchase model made without considering the effect of the product's availability may lead to biased inferences. In the IOS framework, the level of product availability and the number of products already bought is somehow interpreted by customers as a sign of quality product, improving their trust in products. In this regard, Sunil (1988) shows that the number of products already sold influences future purchases. Accordingly, we include these two attributes as learning variables.

2.7. The delivery times

Delivery time has always been a crucial factor in the online purchase framework. Numbers of research pointed out that delivery time is among the factors that substantially influence purchase intention (Nguyen et al., 2018; Rao et al., 2011; Xing, 2010). According to Rao et al. (2011), the delivery delay has an adverse effect on consumer buying conduct. However, Xing (2010) shows that speedy delivery positively influences consumers' purchases. Within the IOS framework, the delivery attribute and its related elements constitute substantial factors that lead to consumer choice during the learning process and purchase. Consequently, we include the delivery time as a learning variable in studying international online consumers' choices and purchases between Chinese brands.

3. Theoretical foundation

3.1. Decision tree algorithm

Most previous studies investigated the effect of learning on purchase and choice, mainly through frameworks and theories associated with the Bayesian model. In line with that, Mehta et al. (2003), Erdem et al. (2004), Iyengar et al. (2007), and Zhao et al. (2011) all utilized the Bayesian framework to model consumer learning using different theories. However, Zhao et al. (2013) employed an extension of the Bayesian framework to model consumer learning, mainly through rating, reviews, and price. Nevertheless, given this research's objective to investigate the effect of the learning variables on choices through ratings and variables related to the products and e-stores, we utilized a machine learning model (decision tree). Supervised learning is a classification technique based on known outcomes. This method has several algorithms (Random forests, decision trees, etc.). The decision tree uses different metrics to determine both the variable to partition and where to create the section. To implement that, we worked with a scikit-learn open-source library through Anaconda. Implementing a Decision tree relies on the Gini method (Gini Impurity and Information Gain) to quantify the level of uncertainty within the dataset. We used entropy reduction to implement the predictive model. The entropy of the impurity D of the training dataset is shown as follows (Gupta et al., 2017):

$$Entropy(D) = \sum_{i=1}^{c} -p \ log_2(p_i) \qquad (3.2.1)$$

p_i is estimated by $|C_{i,D}|/|D|$ to measure the probability that D to class Ci. The evolution of entropy follows the uncertainty inherent in the dataset with a value between 0 to 1

As a classification model, the decision tree is among machine learning techniques (Roy et al., 2019). Among those different classification techniques, the decision tree algorithm is a technique regarded as an efficient method (Elouedi et al., 2000). Accordingly, several studies have been conducted to model consumer behaviour with a decision tree. For instance, with the decision tree, Moore and Carpenter (2010) studied consumers' behaviour toward distributor brands. Costa et al. (2019) utilized a decision tree to discover online reviews' patterns and predict fake evaluations. Naveen et al. (2018) used several classification methods, including a decision tree, to detect user review and rating manipulation. Kumar et al. (2019) combined several machine learning methods to investigate customers' repurchase intentions, including a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2022 doi:10.20944/preprints202208.0148.v1

decision tree. However, despite the forcefulness of different machine learning techniques to examine complicated issues (Moro et al., 2014), the decision tree remains the best for comprehending the result (Costa et al., 2019). Accordingly, we utilize the decision tree in this work.

4. Data

A unique dataset comprising the whole stages of the transaction process has been collected from a Google spreadsheet programming language and Octopus software. We organized the dataset in terms of groups of 3 learning Factors, namely, store factor, product factor, and logistics factor. 44704 transactions from 110 e-retailing stores were analyzed. Table 1 presents a list of the elements and their related variables. The age indicates how long an e-retailer was engaged in international online shopping transactions. That is to say, that variable is taking care of the experiences of e-retailers. The reputation variable tells us the selling behaviour of the e-stores. It is made up of buyers' ratings on product quality and a rating on e-retailers' communication with buyers during the purchase process. We took the average of these two attributes by grouping them under the term reputation since these variables give the buyer information about the selling behaviour and product qualities of the e-retailer. The subscribers' variable of the e-store measures the popularity of the store. For the product factors, we have the quantity of product availability (qpa) and the number of products sold (psold), which takes care of the extent of the demand for an item. We utilized the average product price to implement the model, given the approximate pricing policy used by some e-stores (e.g., $123-150). The Logistics factor is composed of only the Estimated Delivery Time (Edt). Thus, the whole dataset contained eight variables with the target variable (brand). The four brands have within the dataset 3 product lines (e.g., Cover, Screen protector, and Mobile phone) sold to 126 countries from 2017 to 2018. We selected these brands because of their global popularity and within the online selling platform, namely Huawei (Br0=0), Xiaomi (Br1=1), Meizu (Br2=2), and Coolpad (Br3=3). In Table 2, we present the descriptive statistics of variables. Note: In Table 2, the sales are calculated with the number of choices of each brand (transaction) but not with the number of sales exposed by sellers.

5. Results

5.1. Statistical analysis

The learning variables were examined through statistical analysis, correlation matrix, and decision tree to assess their effects on consumer choices and purchases. First, the statistical analysis and correlation matrix have been performed through SPSS. Then we utilized the decision tree algorithm to classify and determine the effects of the learning variables on consumers' choices and purchases. The results offer insights into the role of the learning variables on consumers' choice preferences and options. As shown in Table 2, one observes that all four brands have been sold within e-stores having almost the same average reputation. However, even if the difference is weak, Xiaomi's products are sold in stores with a higher reputation rate in the mean (4.78) compared to its competitors.

Likewise, we see that the average price of Xiaomi's products was higher ($104.04) compared to Huawei ($43.48) and Meizu ($41.01). However, Xiaomi's products were the most purchased (79.91%) compared to Huawei (12.82), Meizu (5.01%), and Coolpad (2.26%). One could say the reputation had influenced sales of Xiaomi products. However, we observe that Coopald products have been sold in almost identical conditions to those of Xiaomi (a high price ($163.58); and a high reputation (4.76)). Nevertheless, Coopald's products are the least sold among brands. In Table 3, we observe that the correlation between reputation and all of the learning variables is significantly negative, except for the correlation between the reputation and the price, which has a positive correlation with prices (0.089; * * p <0.01).

Table.2. Descriptive statistics and average variables per brand

Factor	Variable	Mean	Std. Dev.	Huawei	Xiaomi	Meizu	Coopald
Store's Factor	Age	4.14	2.37	4.11	4.1	3.72	5.55
	Reputation	4.76	0.07	4.74	4.78	4.74	4.76
	Subscribe	156950.3	273040.6	4085	293885	4051	32643

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2022 doi:10.20944/preprints202208.0148.v1

Product Factor	Quantity of Product Availability (qpa)	18626.07	68728.18	43278.76	4211.83	26308.18	4118.76
	Product sold	2783.34	3987.22	1192.43	4260.62	1130.34	1134.53
	Product price	81.43	104.57	43.48	104.04	41.01	163.58
Logistics Factor	Estimated Delivery Time (Edt)	23.69	12.1	27.54	19.9	27.62	29.87
-	brand =choice	0.93	0.8	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	Dataset (%)	-	-	-	-
-	-	Cover	48.54	-	-	-	-
-	-	Screen protector	7.84	-	-	-	-
-	-	Mobile phone	43.66	-	-	-	-
	Brands distribution (%)			43.63	39.1	18.18	0.1
	Sales (%)			12.82	79.91	5.01	2.26

That means the reputation of e-retailers does not significantly influence purchases. However, the positive relationship of the reputation with prices (0.045; * * p <0.01) means that the more the e-retailers have an excellent reputation, the higher the prices of their products are. One may understand these outcomes from the fact that the e-stores utilize these ratings as marketing tools to sell their product at higher prices. However, as shown in this outcome, the rating has no positive influence on choices and purchases. Most online customers, more and more, are aware that these ratings, most of the time, do not reflect the true nature of the e-retailers selling behaviours. The experienced e-buyers are aware that these ratings are manipulated by sellers, as demonstrated by Zhao *et al*. (2013) and Naveen *et al*. (2018). Therefore, purchasers base their purchases and choices on other transactional variables, which are more difficult to manipulate by e-sellers, such as e-retailer selling experience (age), the number of e-retailers' subscribers, the number of products sold, etc. Online selling platforms control the updates of these variables. It is, therefore, almost impossible for retailers to manipulate these variables.

Table 2 shows that Xiaomi, the market leader in sales, has the highest number of subscribers (293885). In Table 3, a positive relationship between subscribers and the number of products sold (.619; **p< .01) as well as between subscribers and price (.460; **p< .01) is highlighted. However, one observes a negative connection with the delivery time (-.531**p< .01).

A positive relationship between number of products sold and the quantity of product availability (.028; ** p< .01) as well as with the price (.205; **p< .01) were found. That means the number of retailers' subscribers and the number of products available for sale influence consumer purchase and choice tendencies within stores having higher prices and shorter delivery times. Accordingly, the price levels, the number of subscribers, and product availability positively influence consumer purchases and choices. This finding could explain why Xiaomi's products have been more purchased than the three other competitors despite their high price levels.

5.2. Decision tree analysis

Transactional data is classified on the basis of the attribute "brand." 70% of the data were randomly utilized to build the decision tree. The result of the decision tree was evaluated with samples' remaining (30%). Table 4 presents the accuracy of the predictive model across each variable. Fig.2 presents the decision tree after implementing the algorithm. The first node is constituted of the attribute or variable of Estimated Delivery Time (Edt). We observe that the variable of Estimated Delivery Time constitutes the first node with the highest entropy value (1.614). Therefore, our four brands have that variable as the basis of classification. That indicates that if one utilized only that single variable (e.g., Edt) to predict the consumers' choices, that variable could only predict with maximum precision of 61.4%.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2022 doi:10.20944/preprints202208.0148.v1

Table.3. Correlation of variables

Variable	1	2	3	4	5	6	7
age	1						
reputation	-.004	1					
subscribers	.339**	-.023**	1				
quantity of product availability	-.002	-.029**	-.136**	1			
product sold	.128**	-.020**	.619**	.028**	1		
price	.045**	.089**	.460**	-.181**	.205**	1	
estimated delivery time (Edt)	-.054**	-.200**	-.531**	.128**	-.432**	.446**	1

Note: ** correlation is significant at the 0.01 level

Fig.2 and Table 4 through the left branch show that 17240 purchases representing about 55.09% of purchases, have been classified and predicted by the algorithm based on only three variables (qpa, price, subscribers) out of 7. Besides, we see that the three most sold brands (Huawei; Xiaomi; Meizu) were chosen based on the price (≤ 67.0 and ≤ 131.5) as a root node. Moreover, we see from Table 3 that 43.93% of Huawei purchases were selected on the basis of the price (≤ 131.5), qpa (≤ 1479.5; qpa ≤ 7374.0; qpa ≤ 9050.0), and subscriber (≤ 113.5; subscriber ≤ 90.5). Therefore, these three learning variables (qpa, price, subscribers) constitute the foundation of consumers' choice preferences and purchases, as already shown by the correlation matrix through the positive relationship of those variables with the quantity of the product sold (e.g., qpa (.028**p< .01); price (.205**p< .01); subscribers (.619; **p< .01)).

To confirm each learning variable's effect on consumers' choice preferences and purchases, we made several simulations to measure the accuracy and misclassified samples. We found that the variables qpa, price, and subscribers are the best variables in terms of prediction when they are considered alone to predict consumers' choice preferences and purchases (Table 4). The prediction rates of these learning variables are respectively 71%, 81%, and 83%. As the subscriber is an attribute related to the store (Table 2), IOCs utilize it to reinforce their trust from the beginning of the transaction process to choose a given e-seller. Then they use the level of price and the quantity of product availability (qpa) to assess product brand originality. That is why, despite the relatively higher costs of the Xiaomi products, they got more sales compared to competitors (Table 2). Xiaomi products fulfilled the three crucial international online shopping requirements, namely the price and the quantity of product availability (qpa) (product originality) and a high subscriber rate (Trust reinforcement). Therefore, one can conclude that international online consumers prefer to buy in stores with a price relatively higher than lower price stores to ensure buying original product brands.

5.3. Discussion, insights, and implications

With this study, we shed new light on the effect of learning variables by providing a theoretical understanding of the influence of the buyers' ratings and the variables related to the transactional components on purchases and choice preferences of Chinese mobile phone brands. Accordingly, the study shows that ratings negatively influence consumers' choice preferences and purchases, unlike previous studies (Zhao et al., 2013; Chevalier and Mayzlin, 2006). However, variables related to the transactional components (e-retailers, product brands, and shipping conditions) are most relevant for international online consumers' preferences and purchases of Chinese mobile phone brands. The study demonstrates that the combination of the three learning variables (qpa, price, subscribers) guides most consumer choice preferences and purchases. The variable of subscribers acts as an influencer of consumer choice of e-retailer outlets through trust reinforcement. Then, given the perceived risk related to this kind of transaction, the price, the quantity of product availability (qpa), and the number of products sold reinforce buyers' trust concerning the product originality. The outcomes indicate that ratings rather than attract consumers to buy at a given e-retailer outlet have repulsive effects on consumers. Consumers believe less in those ratings.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2022 doi:10.20944/preprints202208.0148.v1

Table.4. Classification of consumers' choices based on learning variables

Factors	Variable	Accuracy (%)	Misclassified samples	Variable	Br0's sample (%)	Variables	Br1's sample (%)	Variables	Br2's sample (%)	Variable	Br3's sample (%)
	reputation	60	5348	qpa subscriber price	82.92	Edt reputation Subscriber price	27.72	qpa reputation subscriber price	29.39	Edt	1737(70)
store's factor	age	59	5511							-	-
	subscriber	83	2276							-	-
	pmarket	60	5331							-	-
product factor	psold	57	5766			-	-	-	-	-	-
	color	59	5475			qpa subscriber price	6.6	Edt reputation subscriber price	0.99	-	-
	qpa	71	3893							-	-
	Price	81	2609							-	-
	pricing schemes	56	5921	-	-					-	-
logistics factor	Edt	59	5475	qpa reputation subscriber price	13.19			-			
store's factor	-	76	3236					psold price	0.026	-	-
product factor	-	81	2530							-	-
logistics factor	-	59	5475	-	-	-	-	price	14.92	-	-
store's factor + product factor	-	85	2021	age qpa subscriber price	2.33	age	4.47	-	-	-	-
store's factor, price	-	82	2438					age	14.29	-	-
store's factor + logistics factor	-	74	3508			-	-			-	-
product factor	-	85	2006			psold price	34.92	-	-	-	-
product factor + logistics factor	-	83	2330	-	-			subscriber price	11.63	-	-
price + logistics factor	-	74	3487	-	-	-	-			-	-
Store's factor + product factor	-	82	2385	-	-	-	-			-	-
store's factor + product factor + logistics factor	-	83	2246	-	-	-	-	-	-	-	-
store's factor + price + logistics factor	-	86	1943	Attribute	Br0, Br1, Br2, Sample (%)	-	-	-	-	-	-
product factor +logistics factor +price	-	85	2069	qpa subscriber price	55.09%	-	-	-	-	-	-
store's factor + product factor + price + logistics factor	-	84	2188			-	-	-	-	-	-
-	-	-	-			-	-	-	-	-	-
-	-	-	-			-	-	-	-	-	-

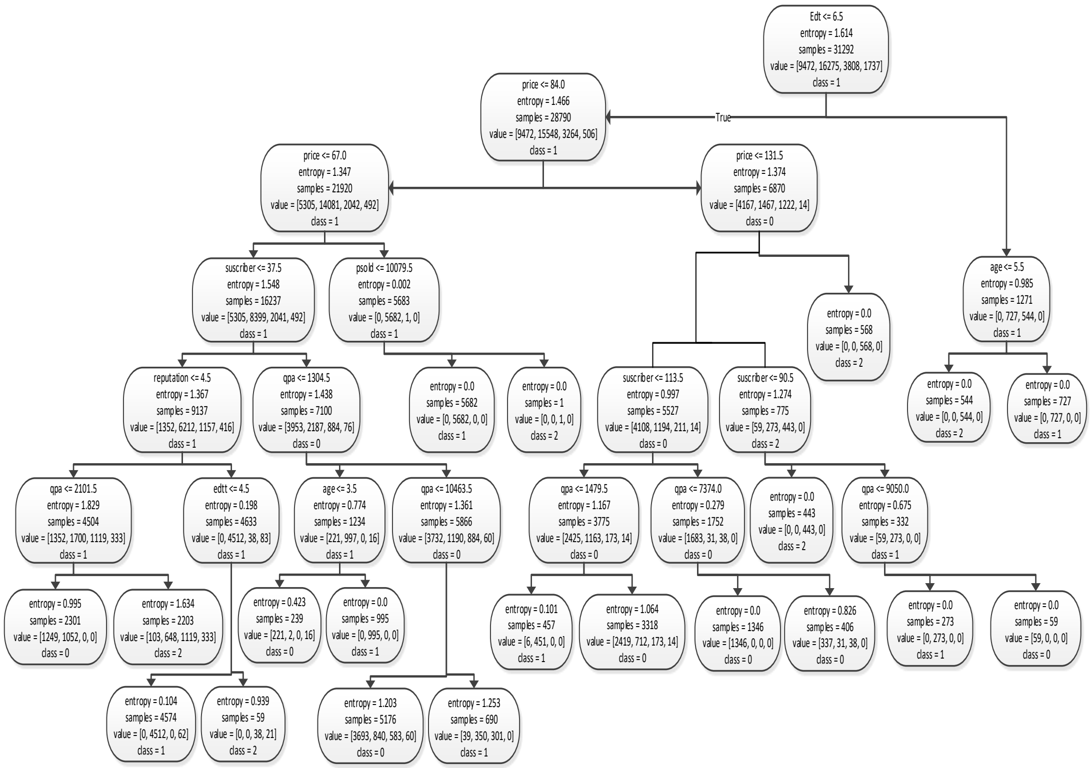


Fig.2. Decision tree

Zhao et al. (2013) show that ratings can increase customers' uncertainty when manipulated. Therefore, their effects on purchase and choice become smaller (Zhao et al., 2013). Accordingly, unlike the study of Chevalier and Mayzlin (2006), which has shown the impact of consumer ratings on choice in the case of book sales, this study indicates that ratings negatively affect consumers' purchases and choice preferences. However, almost all other learning variables related to the transactional components (e-retailers, product brands, and shipping conditions) positively influence consumers' purchases and choices, especially the quantity of product availability (qpa), price, and subscribers. Therefore, based on these results, we can infer that these variables play a substantial role in consumer choice preferences and purchases and are more important than ratings in choice preferences and purchases in the case of international online shopping. As already established by Shah et al. *(*2015) and others, as well as Erdem et al. (2004), this study also confirms respectively that the quantity of product availability (qpa) and the price has a substantial positive impact on consumer purchases and choices. However, this study points out that a high product price level, rather than being negative, positively affects purchases. Besides, one novelty that this study highlight is the effect of subscribers on consumer purchases and choices.

6. Theoretical contribution

First, this study extends the literature on consumer learning by going beyond the learning variables considered in previous studies. It will help the e-stores build a strong marketing strategy according to the effects of those additional learning variables we introduced in the current study. Second, this study highlighted respectively the repulsive and attractive impact of ratings and variables related to the transaction components on consumer choice and purchase. Thirdly, with big data, one observes a rise of machine learning techniques in the studies of consumer behaviours in the marketing field (Chintagunta et al., 2016). However, a fundamental critique is the non-existence of theory in machine learning algorithms (Misra et al., 2019). Using a machine learning model (decision tree), we bring a new approach which associates consumers' behaviour and machine learning to respond to that criticism. Finally, such a theoretical foundation has not yet been developed in consumer learning studies.

Conclusion and limitation

The reinforcement of the consumers' trust constitutes the critical driver of the process of international online consumers' choice and purchases. In line with that, the subscriber attribute is the most crucial element guiding global online consumers in the reinforcement process. Simultaneously, price, the quantity of product availability (qpa), and product sold numbers contribute to assessing the product brand originality. Consequently, e-retailers should consider these variables in elaborating their marketing strategy.

To some degree, this study has certain limitations despite the theoretical and practical contribution that it brings to the online shopping literature. The first limitation of this study concerns the nature of the data. The data is secondary, collected from only one Chinese international online selling platform. Accordingly, further research could be carried out by gathering data from other or several global selling platforms in China or utilizing primary data to investigate. Likewise, future work could incorporate the ratings and other variables related to buyers' experiences, such as reviews, to get a clearer picture of international online consumers' buying strategies. Secondly, researchers could conduct a similar study from other advanced countries in terms of international online selling to verify this study's outcome.

Reference

Arnould, E. J., Price, L., and Zinkhan, G. (2004). *Consumers*. New York: McGraw-Hill.

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, *43*(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345

Chintagunta, P., Hanssens, D. M., & Hauser, J. R. (2016). Editorial—Marketing Science and Big Data. *Marketing Science*, *35*(3), 341–342. https://doi.org/10.1287/mksc.2016.0996

Conlon, C. T., Holland Mortimer, J., Ackerberg, D., Athey, S., Berry, S., Doraszelski, U., Dube, J. P., Haile, P., Hart-Mann, W., Hendricks, K., Imbens, G., Leslie, P., Mortimer, R., Pakes, A., & Wolak, F. (2012). *Demand Estimation Under Incomplete Product Availability*. https://pdfs.semanticscholar.org/ee45/bac4263cf63baa9d12702b644b6c4b3ff730.pdf

Constantinides, E. (2004). Influencing the online consumer's behavior: The Web experience. In *Internet Research*. https://doi.org/10.1108/10662240410530835

Costa, A., Guerreiro, J., Moro, S., & Henriques, R. (2019). Unfolding the characteristics of incentivized online reviews. *Journal of Retailing and Consumer Services*, *47*(April 2018), 272–281. https://doi.org/10.1016/j.jretconser.2018.12.006

Elouedi, Z., Mellouli, K., & Smets, P. (2000). Classification with Belief Decision Trees in: ArtificialIntelligence:Methodology,Systems,andApplications. 80–90. https://doi.org/10.1007/3-540-45331-8_8

Fang, F., & Xu, X. (2011). An analysis of consumer training for feature rich products. *Decision Support Systems*, *52*(1), 169–177. https://doi.org/10.1016/J.DSS.2011.07.001

Gupta, B., Uttarakhand, P., & Rawat, I. A. (2017). Analysis of Various Decision Tree Algorithms for Classification in Data Mining. *International Journal of Computer Applications*, *163*(8), 975–8887. https://pdfs.semanticscholar.org/fd39/e1fa85e5b3fd2b0d000230f6f8bc9dc694ae.pdf

Hong, H., Xu, D., Xu, D., Wang, G. A., & Fan, W. (2017). An empirical study on the impact of online word-of-mouth sources on retail sales. *Information Discovery and Delivery*, *45*(1), 30–35. https://doi.org/10.1108/IDD-11-2016-0039

Iyengar, R., Ansari, A., & Gupta, S. (2007). A Model of Consumer Learning for Service Quality and Usage. *Journal of Marketing Research*, *44*(4), 529–544. https://doi.org/10.1509/jmkr.44.4.529

Kostyra, D. S., Reiner, J., Natter, M., & Klapper, D. (2016). Decomposing the effects of online customer reviews on brand, price, and product attributes. *International Journal of Research in Marketing*, *33*(1), 11–26. https://doi.org/10.1016/j.ijresmar.2014.12.004

Kumar, A., Kabra, G., Mussada, E. K., Dash, M. K., Rana, P. S., Hu, Y., Chen, C., Costa, A., Guerreiro, J., Moro, S., & Henriques, R. (2019). Combined artificial bee colony algorithm and machine learning techniques for prediction of online consumer repurchase intention. *Neural Computing and Applications*, *31*(April 2018), 877–890. https://doi.org/10.1007/s00521-017-3047-z

Lakshmanan, A., & Krishnan, H. S. (2011). The Aha! Experience: Insight and Discontinuous Learning in Product Usage. *Journal of Marketing*, *75*(6), 105–123. https://doi.org/10.1509/jm.10.0348

Lin, A. J., Li, E. Y., & Lee, S. Y. (2018). Dysfunctional customer behavior in cross-border e-commerce: A Justice-affect-behavior model. *Journal of Electronic Commerce Research*, *19*(1), 36–54.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, *70*(3), 74–89. https://doi.org/10.1509/jmkg.70.3.74

Mehta, N., Rajiv, S., & Srinivasan, K. (2003). Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation. *Marketing Science*, *22*(1), 58–84. https://doi.org/10.1287/mksc.22.1.58.12849

Misra, K., Schwartz, E. M., & Abernethy, J. (2019). Dynamic Online Pricing with Incomplete Information Using Multiarmed Bandit Experiments. *Marketing Science*, *38*(2), 226–252. https://doi.org/10.1287/mksc.2018.1129

Mo, Z., Li, Y.-F., & Fan, P. (2015). Effect of Online Reviews on Consumer Purchase Behavior. *Journal of Service Science and Management*, *08*(03), 419–424. https://doi.org/10.4236/jssm.2015.83043

Moore, M., & Carpenter, J. M. (2010). A decision tree approach to modeling the private label apparel consumer. *Marketing Intelligence and Planning*, *28*(1), 59–69. https://doi.org/10.1108/02634501011014615

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*, 22–31. https://doi.org/10.1016/J.DSS.2014.03.001

Naveen Kumar, Deepak Venugopal, L. Q., & Kumar, S. (2018). Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Leraning. *Journal of Management Information Systems*, 350–380. https://doi.org/10.1080/07421222.2018.1440758

Nguyen, D. H., de Leeuw, S., & Dullaert, W. E. H. (2018). Consumer Behaviour and Order Fulfilment in Online Retailing: A Systematic Review. In *International Journal of Management Reviews* (Vol. 20, Issue 2, pp. 255–276). https://doi.org/10.1111/ijmr.12129

Rao, S., Griffis, S. E., & Goldsby, T. J. (2011). Failure to deliver? Linking online order fulfillment glitches with future purchase behavior. *Journal of Operations Management*, *29*(7–8), 692–703. https://doi.org/10.1016/j.jom.2011.04.001

Roy, S., Mondal, S., Ekbal, A., & Desarkar, M. S. (2019). Dispersion Ratio based Decision Tree Model for Classification. *Expert Systems with Applications*, *116*, 1–9. https://doi.org/10.1016/J.ESWA.2018.08.039

Shah, D., Kumar, V., & Zhao, Y. (2015). Diagnosing Brand Performance: Accounting for the Dynamic Impact of Product Availability with Aggregate Data. *Journal of Marketing Research*, *52*(2), 147–165. https://doi.org/10.1509/jmr.13.0530

Sunil Gupta. (1988). Impact of Sales Promotion on When, What, and How Much To Buy. *Journal of Marketing Research*, *25*(4), 342–355.

Wang, X., Xu, Y., Wang, L., Xu, X., & Chen, Y. (2017). Transmission of information about consumer product quality and safety: a social media perspective. *Information Discovery and Delivery*, *45*(1), 10–20. https://doi.org/10.1108/IDD-10-2016-0035

Wang, Y., Wang, Y., & Lee, S. H. (2017). The effect of cross-border e-commerce on China's international trade: An empirical study based on transaction cost analysis. *Sustainability (Switzerland)*, *9*(11), 1–13. https://doi.org/10.3390/su9112028

Xing, Y., Grant, D. B., McKinnon, A. C., & Fernie, J. (2010). Physical distribution service quality in online retailing. *International Journal of Physical Distribution & Logistics Management*, *40*(5), 415–432. https://doi.org/10.1108/09600031011052859

Yrj, M., & Saarij, H. (2019). The Seller ' s Experience in Consumer-to-Consumer E-Commerce. *Springer*

Nature Switzerland AG, *Leonard 2011*, 145–153. https://doi.org/10.1007/978-3-030-18911-2

Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling Consumer Learning from Online Product Reviews. *Marketing Science*, *32*(1), 153–169. https://doi.org/10.1287/mksc.1120.0755

Zhao, Y., Zhao, Y., & Helsen, K. (2011). Consumer learning in a turbulent Market environment: Modeling Consumer Choice Dynamics after a Product-harm Crisis. *Journal of Marketing Research*, *48*(2), 255–267. https://doi.org/10.1509/jmkr.48.2.255

Zhu, F., & Zhang, X. (Michael). (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, *74*(2), 133–148. https://doi.org/10.1509/jmkg.74.2.133