Exploration on Consumers' Cognition, Coping Strategies and Future Consumption Intentions Regarding Price Discrimination in the Big Data Era

Yanran Cao[1, a], Dongjing Chen[2, *]

[1]Campion School, Athens, Greece

[2]AmazingX Academy, Guangdong, China

[a]yanrancao07@gmail.com, * Corresponding Author Email: sysucdj@amazingx.org

Abstract. In the context of the big data era, consumer rights protection has become a crucial element for the healthy development of the digital economy. This study investigates the perception and coping strategies of consumers in the Xinjiang Uyghur Autonomous Region regarding the phenomenon of big data "price gouging." By employing descriptive statistical analysis, Apriori association rule mining, and the C4.5 decision tree model, the study systematically analyzes the differences in consumer perception of big data "price gouging," the influencing factors of coping strategies, and the mechanisms affecting consumption intentions. The findings reveal that there are significant differences in consumer perception based on demographic groups, with older, lower-income, and less-educated consumers having relatively lower levels of awareness, while younger, highly educated, and higher-income groups have higher levels of awareness. Consumers' coping strategies mainly focus on improving information literacy, platform selection, and adjusting price sensitivity, among which information literacy and consumer psychological factors have the most significant impact on purchase intentions. Consumers generally pay close attention to fairness, transparency, and information security in the consumption process, and there is a need to further optimize relevant policies and regulations to strengthen rights protection. The C4.5 decision tree model simulation indicates that consumers over the age of 25 with higher online shopping frequency are more likely to be affected by big data "price gouging." This study provides empirical evidence for the formulation of consumer rights protection policies in the digital economy and has important reference value for promoting the healthy development of the digital economy.

Keywords: Price Discrimination; Big Data "Price Gouging"; C4.5 Decision Tree Model; Apriori Association Rules.

1. Introduction

In the context of the digital economy, big data technology has become a core driver for business model innovation. With the rapid development of artificial intelligence, mobile payments, and other technologies, the interaction between enterprises and consumers has become increasingly frequent, laying a solid foundation for the big data economy. However, behind this technological dividend lies a growing social issue—the phenomenon of big data "price gouging." According to a survey report by the Beijing Consumer Association, 88.32% of respondents believe that big data "price gouging" is widespread, and 56.92% have experienced it.

Big data price gouging is essentially a differential pricing strategy based on user profiling. Platform enterprises collect and analyze consumers' personal information, browsing records, and consumption habits to implement price discrimination for different types of users[1]. This phenomenon is particularly common in online travel, e-commerce platforms, and ride-hailing services, manifesting as different prices for the same product or service for different users, often with higher prices for long-term users than new users and higher prices for Apple users than Android users[2]. This business practice not only infringes on consumer rights but also undermines the fair competition environment of the market economy, potentially affecting the healthy and orderly development of China's digital economy in the long run.

Existing research has explored the phenomenon of big data price gouging from multiple perspectives. In terms of definition, most scholars regard it as a form of price discrimination[3], although some researchers argue that non-price discrimination behaviors may also lead to price gouging[4]. Regarding the motivations for price gouging, some researchers believe that high profits are the fundamental reason[5], while others explore the balance between price gouging and long-term consumer relationships from the perspective of customer loyalty[6].

In terms of consumer rights protection, existing studies generally agree that big data price gouging violates consumers' rights to be informed and to fair transactions[7]. Some studies point out that big data price gouging often targets short-term transactions with consumers, without fully considering long-term relationships[8]. However, existing research mostly focuses on legal regulation and industry norms, with relatively few studies on consumer perception and long-term consumption intentions, especially those lacking in-depth analysis based on large-scale empirical data.

This study employs descriptive statistical analysis, Apriori association rules, and the C4.5 decision tree model to investigate consumers' perception of price discrimination, coping strategies, and the impact mechanisms on future consumption intentions in the big data era, using residents from five cities and prefectures in the Xinjiang Uyghur Autonomous Region as research subjects. The study fills the gap in existing literature regarding consumer perception of price discrimination in emerging markets, provides empirical evidence for the association between consumer characteristics and coping strategies, and reveals the core impact of price discrimination on consumption intentions through quantitative analysis, offering a new research perspective for consumer behavior theory and regulatory policies in the digital economy.

2. Research Methods and Model Construction

2.1. Data Acquisition and Validation

The questionnaire designed for this study primarily focuses on consumers' basic information (gender, age, education, occupation, income, etc.), their perception of big data price gouging, experiences with big data price gouging, coping strategies, and future consumption intentions. A total of 1,523 questionnaires were distributed to residents in five cities and prefectures in the Xinjiang Uyghur Autonomous Region, including Kashgar and Urumqi, with 1,376 valid questionnaires recovered, resulting in an effective recovery rate of 90.35%. To ensure the validity of the questionnaire data, Cronbach's α coefficient was used for reliability testing.

$$\alpha = \frac{k}{k-1}\left(1 - \sum_{i=1}^{k} \sigma_i^2 / \sigma_{\mathrm{T}}^2\right) \tag{1}$$

where k is the number of questions in the questionnaire, σ_i^2 is the variance of the score of the i -th question, and σ_{T}^2 is the variance of the total score of the questionnaire.

After calculation, the Cronbach's α coefficient of this questionnaire is 0.91, indicating that the questionnaire has high reliability.

2.2. Apriori Association Rules

Association analysis, which involves discovering frequently occurring itemsets in a given dataset, identifies regularities between two or more variables. Data association is an important type of knowledge that can be discovered in databases. The algorithm used for association rule analysis in this study is the Apriori algorithm, with the following basic definitions:

Let I represent the set of all items, and D represent the transaction database. An element in a transaction T is called an item; a set of items is called an itemset; an itemset containing k items is referred to as a k-itemset.

Definition 1: Support of an itemset. The support of an itemset A is defined as the proportion of transactions in the transaction database D that contain A. That is,

$$\text{support}(A) = \frac{\text{number of transactions containing A}}{\text{total number of transactions in D}} \tag{2}$$

Definition 2: Frequent itemset. If the support of an itemset A is not less than a predefined minimum support threshold (min_sup), i.e., support $(A) \geq \text{min\_sup}$, then itemset A is considered a frequent itemset. A frequent itemset containing k items is called a frequent k-itemset.

Definition 3: Association rule. An association rule can be represented as a logical implication $A \Rightarrow B$, where A and B are two distinct non-empty subsets of I. The strength of an association rule is typically measured by support, confidence, and lift.

Definition 4: Support of an association rule. The support of an association rule $A \Rightarrow B$ is defined as the proportion of transactions in the transaction database D that contain $A \cup B$. That is,

$$\text{support}(A \Rightarrow B) = \frac{\text{number of transactions containing A} \cup \text{B}}{\text{total number of transactions in D}} \tag{3}$$

Definition 5: Confidence of an association rule. The confidence of an association rule $A \Rightarrow B$ is defined as the proportion of transactions containing A that also contain B. That is,

$$\text{confidence}(A \Rightarrow B) = \frac{\text{support\_count}\ (A \cup B)}{\text{support\_count}\ (A)} \tag{4}$$

where support_count (A) and support_count $(A \cup B)$ are the number of transactions in the transaction database D that contain A and $A \cup B$, respectively.

Definition 6: Lift of an association rule. In addition to support and confidence, lift is another important criterion for evaluating association rules. The lift of an association rule $A \Rightarrow B$ is defined as the ratio of the proportion of transactions containing $A \cup B$ to the proportion of transactions containing B. That is,

$$\text{lift}(A \Rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A) \times \text{support}(B)} \tag{5}$$

If $\text{lift}(A \Rightarrow B) < 1$, it indicates a negative association between itemsets A and B, meaning the presence of A suppresses the occurrence of B; if $\text{lift}(A \Rightarrow B) = 1$, it indicates that A and B are independent and unrelated; if $\text{lift}(A \Rightarrow B) > 1$, it indicates a positive association between A and B, meaning the presence of A promotes the occurrence of B, and the greater the lift value, the stronger the promoting effect.

In this study, the factors influencing big data price gouging were normalized into eight internal and external factors: price fairness, platform reliability, probability of choosing price gouging, consumer information assessment, consumer outcome reflection, evaluation of purchase decisions, product differentiation on purchase decisions, and government policies to avoid price gouging. To deeply analyze the internal associations between normalized big data price gouging and these eight factors, the sample answers to the question "Which factors do you think can regulate big data price gouging?" from authoritative literature were used as reference values to comprehensively analyze the associations between these eight factors and the regulation of big data price gouging.

2.3. C4.5 Decision Tree Model

The C4.5 decision tree model is an effective method for studying the association between purchase factors and the probability of big data price gouging. It helps reveal the potential relationships between purchasing behavior and big data price gouging. C4.5 is a classic decision tree learning algorithm used to generate decision tree models from data. It recursively partitions the dataset, selecting the best splitting attribute to construct the decision tree model, which can be used for classification and regression tasks.

The core metric of the C4.5 algorithm is the gain ratio, which is used to select the optimal splitting attribute. The gain ratio is calculated as follows:

$$\mathrm{GainRatio(A)} = \frac{\mathrm{Gain(A)}}{\mathrm{SplitInfo(A)}} \tag{6}$$

where $\mathrm{Gain(A)}$ is the information gain of attribute A for the dataset D, calculated as:

$$\mathrm{Gain(A)} = \mathrm{Entropy(D)} - \sum_{\mathrm{v \in Values(A)}} \frac{|\mathrm{D_v}|}{|\mathrm{D}|} \mathrm{Entropy(D_v)} \tag{7}$$

The entropy of the dataset D is calculated as:

$$\mathrm{Entropy(D)} = -\sum_{\mathrm{i=1}}^{\mathrm{c}} \mathrm{p_i} \log_2 \mathrm{p_i} \tag{8}$$

where c is the number of classes in the dataset D, and $\mathrm{p_i}$ is the proportion of samples of class i in the dataset D.

The split information of attribute A for the dataset D is calculated as:

$$\mathrm{SplitInfo(A)} = -\sum_{\mathrm{v \in Values(A)}} \frac{|\mathrm{D_v}|}{|\mathrm{D}|} \log_2 \frac{|\mathrm{D_v}|}{|\mathrm{D}|} \tag{9}$$

where $\mathrm{D_v}$ is the subset of D where attribute A takes the value v.

Based on the above calculations of the gain ratio, the C4.5 decision tree model recursively selects the optimal splitting attribute to construct the decision tree. At each split, the attribute with the highest gain ratio is chosen as the splitting attribute until the stopping criteria are met (e.g., all samples in the dataset belong to the same class, or the gain ratio is less than a threshold).

3. Results

Descriptive statistical analysis primarily serves to describe the distribution and categorical relationships of survey variables. In this study, it was used to describe the demographic characteristics and online shopping behaviors of the respondents, analyzing the proportions or distributions of different groups within the sample.

From Table.1, it can be observed that the majority of respondents have a basic understanding of big data price gouging (41.28%), followed by those who have a somewhat limited understanding (22.82%). Only a small proportion of respondents have a very good understanding (18.68%), while those who have little understanding are slightly fewer than those who have a very good understanding (14.32%). The number of respondents who have never heard of the concept is the smallest (2.91%). Among the respondents, 76.24% have experienced big data price gouging, while 23.76% have not. This indicates that the majority of respondents have encountered big data price gouging in their consumption experiences. Regarding the impact of big data price gouging avoidance applications on consumer interests, 84.45% believe that these applications are beneficial and have a significant impact, while only a small proportion think they are useless (7.05%), have no impact (5.45%), or are not familiar with them (3.05%). This suggests that most people believe that big data price gouging avoidance applications are effective in protecting consumer rights and providing a fairer consumption environment.

Table 1. Descriptive Statistics of Respondents' Understanding of Big Data Price Gouging

Characteristic Variable	Type	Sample Count	Percentage (%)	Cumulative Percentage (%)
Understanding of Big Data Price Gouging	Very Well Understood	257	18.68	18.68
	Basically Understood	568	41.28	59.96
	Somewhat Understood	314	22.82	82.78
	Basically Not Understood	197	14.32	97.09
	Never Heard of It	40	2.91	100
Experience with Big Data Price Gouging	Yes	1049	76.24	76.24
	No	327	23.76	100
Impact of Big Data Price Gouging on Consumer Fairness	Strongly Agree	513	37.28	37.28
	Agree	570	41.42	78.71
	Neutral	214	15.55	94.26
	Disagree	20	1.45	95.71
	Strongly Disagree	59	4.29	100
Effectiveness of Big Data Price Gouging Avoidance Applications	Significant Impact	1162	84.45	84.45
	No Use at All	97	7.05	91.5
	No Impact	75	5.45	96.95
	Don't Know	42	3.05	100

Overall, most people have some understanding of big data price gouging, and the majority believe that applications to avoid big data price gouging are beneficial and have a significant impact. However, there is still a portion of people who have limited understanding or hold different views, reflecting the diverse concerns and perspectives of consumers regarding data privacy and personal rights in the big data era.

Table 2. Basic Demand Association Rules between Normalized Big Data Price Gouging and Various Factors

Rule Number	Rule	Instance Count	Rule Confidence	Rule Support	Lift
R1	Online Reputation, Consumer Psychological Factors ⇒ Platform Factors	77	80.82%	10.13%	3.28
R2	Price Sensitivity, Consumer Psychological Factors ⇒ Information Literacy	83	80.00%	13.87%	3.264
R3	Online Reputation, Consumer Psychological Factors ⇒ Information Literacy	73	79.05%	14.56%	3.035
R4	Online Reputation, Consumer Psychological Factors ⇒ Price Sensitivity	77	78.50%	12.90%	2.926
R5	Purchase Decision Influence, Price Sensitivity ⇒ Information Literacy	77	77.92%	10.68%	2.92
R6	Purchase Decision Influence, Consumer Psychological Factors ⇒ Information Literacy	76	76.25%	10.54%	2.866
R7	Platform Factors, Information Literacy ⇒ Consumer Psychological Factors	80	76.19%	11.10%	2.863
R8	Purchase Decision Influence, Consumer Psychological Factors ⇒ Price Sensitivity	105	75.31%	14.56%	2.861
R9	Purchase Decision Influence, Price Sensitivity ⇒ Consumer Psychological Factors	100	75.00%	11.23%	2.859

The association rules derived from the Apriori algorithm are shown in Table.2. The actions or thoughts of respondents in the face of related scenarios can be analyzed to understand their focus points of attention and demand. The focus points of attention refer to the characteristics or attributes that users pay the most attention to during continuous consumption, while the focus points of demand refer to the needs and desires of users for normalizing big data price gouging during continuous consumption. The following are analyses of these focus points:

3.1. Users' Focus Points of Attention

Online Reputation: This refers to consumer evaluations and comments on products on the internet. From the association rule table, it can be seen that users often consider the online reputation of products when deciding on their information literacy and price sensitivity. Therefore, relevant platforms need to pay attention to their brand image and reputation, actively regulate big data price gouging behavior, provide users with a fair, transparent, and trustworthy consumption experience, and promptly respond to user feedback and complaints to enhance user trust and loyalty.

Platform Factors: These refer to the influence of the platform where products are purchased on users. From the association rule table, it can be seen that users may prefer to purchase products on well-known e-commerce platforms, as these platforms typically offer more choices and better services. Therefore, companies need to consider how to improve their platform competitiveness while ensuring consumer rights to attract more users and consumers.

Price Sensitivity: This refers to the degree to which users are sensitive to changes in product prices, including price fluctuations that may occur due to big data price gouging (e.g., different prices for different mobile device brands and price increases after multiple views of the same product). From the association rule table, it can be seen that users' price sensitivity can have a certain impact on consumer psychological factors (trust in the platform). Therefore, companies need to price reasonably and avoid "double standards" in big data price gouging behavior to ensure consumer rights and win user trust.

3.2. Users' Focus Points of Demand

Information Literacy: This refers to users' proficiency in information technology and their ability to comprehensively assess and apply information. From the association rule table, it can be seen that consumers' information literacy is often influenced by personal preferences such as purchase decision impact, price sensitivity, and consumer psychological factors. Information literacy is an important means for individuals to avoid and review big data price gouging. Users need to continuously improve their purchase decision impact, price sensitivity, and consumer psychological factors to enhance their ability to discern information, effectively use technical information, assess the relevance and credibility of information, and evaluate the process and results of information use. Only in this way can users avoid being gouged by big data in consumption.

Consumer Psychological Factors: These include consumers' values, attitudes, beliefs, and emotions towards platforms. These factors can influence users' purchasing behavior and preferences. From the above association rule table, it can also be seen that consumer psychological factors are very important. Therefore, in the face of the possibility of big data price gouging, we need not only regulatory authorities to adjust but also consumers to regulate their psychological factors to prevent big data from grasping their direct purchasing behavior preferences.

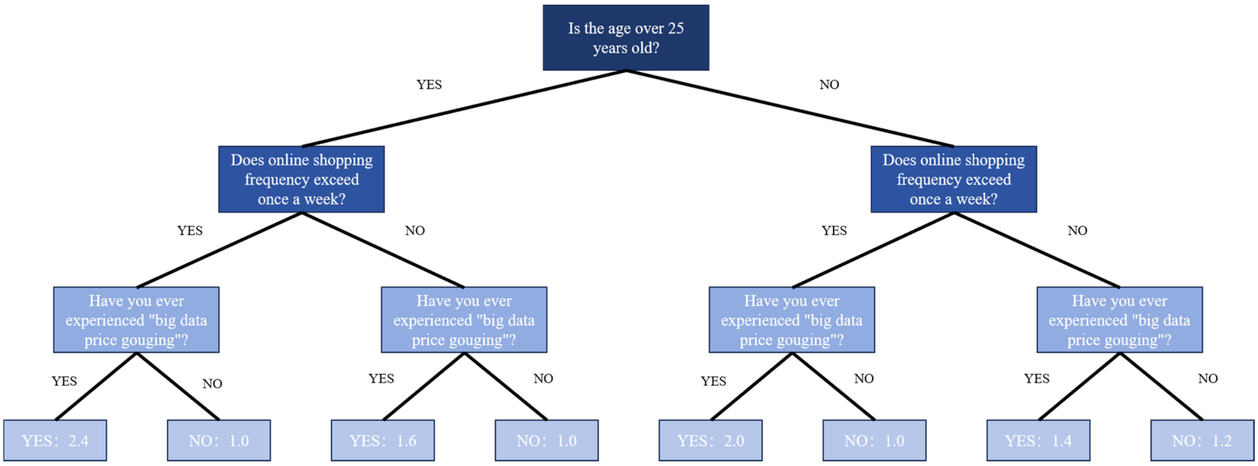

Figure 1. Ecological Benefits of Green Buildings

As shown in Figure 1, the decision tree generated from the 1,376 sample data set reveals the following: (1) In the decision tree, "whether age is over 25" and "whether online shopping frequency exceeds once a week" have good distinguishing power and play an important role in whether consumers are subject to big data price gouging. (2) From the decision tree, it can be seen that consumers with "online shopping frequency exceeding once a week" have more frequent contact with online shopping platforms and product information, thus increasing the likelihood of being affected by big data price gouging. Consumers with high online shopping frequency may be more susceptible to personalized pricing strategies, leading to a higher probability of big data price gouging. (3) For the test sample ["yes", "yes", "yes"], the classification result is {"yes": 2.4, "no": 1.0}, so the classification of this sample is "yes." This means that consumers who are "over 25 years old" and have "online shopping frequency exceeding once a week" are more likely to be subject to big data price gouging.

By analyzing these decision nodes, a deeper understanding of the correlation between purchase factors and the probability of big data price gouging can be achieved. This helps to better understand consumer behavior patterns, effectively cope with the phenomenon of big data price gouging, and improve consumer satisfaction. Based on the experimental results, the following conclusions can be drawn: Consumers over the age of 25 typically have more consumption history data available for analysis. These data can help companies more accurately understand consumers' preferences and behavior habits, thereby implementing personalized pricing and marketing strategies, including big data price gouging. Consumers over the age of 25 may have higher consumption capacity and demand, making them more likely to become targets for companies' personalized pricing strategies. Frequent online shoppers typically leave more data traces on platforms, enabling companies to better analyze their behavior and implement big data price gouging strategies. Frequent online shoppers may be considered a more active and sensitive consumer group, making them more likely to become targets for big data price gouging strategies.

4. Conclusions

This study focuses on investigating consumers' perception of price discrimination, coping strategies, and future consumption intentions in the big data era. By analyzing 1,376 valid questionnaires from five cities and prefectures in Xinjiang and employing methods such as Apriori association rules and the C4.5 decision tree model, the following conclusions were drawn: In terms of consumer perception, the majority of consumers have some understanding of big data price gouging, with over 70% having experienced it. Most consumers also believe that applications to avoid price gouging are significantly beneficial for protecting their rights. However, there are still some consumers who have limited understanding or hold different views. Regarding influencing factors, online reputation, platform factors, and price sensitivity are the focus points of consumer attention, while information literacy and consumer psychological factors are the focus points of demand. There are complex associations between these factors. The decision tree analysis indicates that consumers over the age of 25 with online shopping frequency exceeding once a week are more likely to be subject to big data price gouging. This is because such consumers have more consumption data available for analysis, higher consumption capacity and demand, and leave more data traces on platforms. Overall, the phenomenon of big data price gouging is widespread, causing damage to consumer rights and the market economy environment. Consumers need to improve their information literacy and psychological regulation abilities. Relevant platforms should standardize their behavior, and the government should strengthen the implementation of avoidance policies to promote the healthy development of the digital economy.

References

[1] Sinha A, Gupta R, Kutnar A. Sustainable development and green buildings[J]. Drvna industrija, 2013, 64(1): 45-53.

[2] Howe J C. Overview of green buildings[J]. Envtl. L. Rep. News & Analysis, 2011, 41: 10043.

[3] Dwaikat L N, Ali K N. Green buildings cost premium: A review of empirical evidence[J]. Energy and Buildings, 2016, 110: 396-403.

[4] Shen Q, Chen Q, Tang B, et al. A system dynamics model for the sustainable land use planning and development[J]. Habitat international, 2009, 33(1): 15-25.

[5] Zhang X, Wang C, Li E, et al. Assessment model of ecoenvironmental vulnerability based on improved entropy weight method[J]. The Scientific World Journal, 2014, 2014(1): 797814.

[6] Pan B, Liu S, Xie Z, et al. Evaluating operational features of three unconventional intersections under heavy traffic based on CRITIC method[J]. Sustainability, 2021, 13(8): 4098.

[7] Ouyang Y. Evaluation of river water quality monitoring stations by principal component analysis[J]. Water research, 2005, 39(12): 2621-2635.

[8] Sun Y, Liang X, Xiao C. Assessing the influence of land use on groundwater pollution based on coefficient of variation weight method: A case study of Shuangliao City[J]. Environmental science and pollution research, 2019, 26(34): 34964-34976.