Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 4
EMSD 2023

Analysis of Factors Influencing Consumers' Satisfaction with Online Fresh Food Shopping

Xinyu Gao

College of Economics and Management, Beijing Jiaotong University, Beijing, China

gaoxinyu991119@163.com

Abstract. With the development of Internet technology, more and more people choose online shopping, resulting in an increasing amount of online review data. It is crucial to mine information about users' feelings from these reviews to obtain factors that affect customer satisfaction. This article uses crawler technology to obtain the review data of JD supermarket hairy crabs. After data preprocessing, it extracts five characteristic factors that affect satisfaction, including logistics, specifications, taste, freshness and quality. Then, it uses a Likert scale to quantify and score the data, and establishes a customer satisfaction factor model based on Bayesian network, which is compared with a neural network model. Evaluation criteria include accuracy, precision, and recall. The research results show that the customer satisfaction model based on Bayesian network has the best measurement effect, which can output the influence degree and correlation of each factor on satisfaction, and provide relevant suggestions for producers and other stakeholders.

Keywords: Online Reviews; Customer Satisfaction; Bayesian Network Model.

1. Introduction

On August 28, 2023, the China Internet Network Information Center (CNNIC) released the 52nd "Statistical Report on the Development of China's Internet Network" in Beijing. The report shows that as of June 2023, the number of Internet users in China reached 1.079 billion, an increase of 11.09 million compared to December 2022, and the Internet penetration rate reached 76.4%. The data in the report show that the developed internet allows people to actively spread and share information instead of passively accepting it. With the development and prosperity of the Internet, people are also changing their lifestyles in keeping with the times, among which online shopping brings great convenience to people.

Online shopping malls not only provide consumers with a variety of products and shopping convenience, but also provide a platform for consumers to share and communicate. People choose the goods they want to buy through e-commerce platforms. After users receive the goods and try them out for a certain period of time, they will have an intuitive evaluation of the quality of the products or the service of the e-commerce platform. When users post their real feelings in the comments of the e-commerce website, online reviews are generated. It is very important to explore the users' concerns based on the content of the comments, and how these concerns affect user satisfaction. This article studies customer satisfaction based on the online reviews of hairy crabs on JD.com, and conducts in-depth analysis of online evaluation data through data mining technology. Firstly, Octopus will be used for data collection to obtain the required data from the shopping reviews of JD.com. Secondly, data cleaning technology will be used to ensure the quality and consistency of the data. Then, IBM SPSS Modeler will be used for data analysis and mining to extract valuable information. Finally, the results will be visualized and the analysis results will be explained and analyzed through professional knowledge.

2. Data Introduction

2.1. Data Collection and Cleaning

This study takes the hairy crabs from JD.com as the research object, and the main purpose of crawling data is to obtain relevant information about product sales and online reviews. The amount of data collected in this study is 1200 pieces, and the data set includes the following categories: customer ID, level, comment content, comment time, evaluation score. During the collection process, duplicate parts are automatically filtered, and the results are imported into Excel.

The raw data collected by the Octopus collector cannot be directly analyzed. It requires data cleaning and screening to ensure data accuracy before subsequent problem induction and deep data analysis. Data screening refers to the process of performing data cleaning, consistency checking, and removing invalid and missing values from customer online reviews.

According to the text content in the, the following steps were taken in order: first, the collected 1200 pieces of data were saved and copied, and operations were carried out under the premise of ensuring the integrity of the data source was not destroyed; second, the "comment content" column was screened to remove missing values and invalid comments, resulting in 1076 pieces of data.

2.2. Data Preprocessing

High-frequency words generally refer to words that appear frequently and are useful in documents. High-frequency words in online reviews are words that appear frequently in reviews and express consumers' attitudes towards products. In this article, the extraction of high-frequency words mainly utilizes a Chinese word segmenter to segment words, and then counts the number of occurrences of each word.

From the sorting results of the data, consumers have more concerns about the logistics, specifications, taste, and quality of hairy crabs. The subject terms with similar meanings or describing the same subject are merged into new factors, resulting in six variables, namely logistics, specifications, taste, freshness, and quality. These are the attributes of products and customer service that consumers focus on. Using these five factors to construct a customer satisfaction model, the word frequency statistics of the characteristic factors are summarized as shown in Table 1.

Table 1. Summary Table of Feature Factor Word Frequency

Feature Factors	Frequency
logistics	458
specifications	308
taste	392
freshness	468
quality	286

As shown in the table above, customers' concerns about hairy crabs are reflected in the frequency of each factor. The higher the frequency, the higher the customer's level of attention to that factor. The five variables obtained above, to some extent, reflect consumers' consumption tendencies. However, these data are unstructured, so for subsequent analysis, we will quantize and score the data using a five-point scale, which is commonly used in questionnaire surveys. By comparing with the Likert scale, we establish a scoring standard.

Table 2. Scoring Standard Table

Scoring Standard	Score
strongly satisfied	5
satisfied	4
neutral	3
unsatisfied	2
strongly satisfied	1

In online comments, there may be certain words like "very", "extremely", "especially" that modify the characteristics of products. By analyzing these modifying words, we can see customers' satisfaction levels for products. Therefore, we establish a customer satisfaction sentiment lexicon and scoring standard based on the modifying words of each feature factor, and use this to quantize and score the feature factors.

During the scoring process, scores are given based on the emotional tendency of the modifying words. When scoring similar words, it is important to consider synonym conversion. For feature factors that are not mentioned in the comment data, in this article, a consistent score of 3 is given, indicating a neutral emotional tendency. The scoring process strictly follows the scoring rules, and the resulting data is used as the basic data for constructing customer satisfaction. At the same time, in crawling online comment data from JD.com, overall ratings from consumers' comments are collected to represent consumers' overall satisfaction level for this purchase.

3. Construction of the Model of Influencing Factors of Customer Satisfaction

3.1. Model Construction of Influencing Factors of Customer Satisfaction

Firstly, import the processed data into IBM SPSS Modeler software, and establish a Bayesian network model of customer satisfaction factors through type selection and other operations. Then, analyze the importance of the factors affecting customer satisfaction and the relationship between the factors, and evaluate the accuracy of the model. IBM SPSS Modeler has a variety of data mining algorithms, with simplified operation procedures and intuitive and easy-to-understand analysis results. It is a powerful data mining software. The process of Bayesian modeling using IBM SPSS Modeler software is as shown in Figure 1.

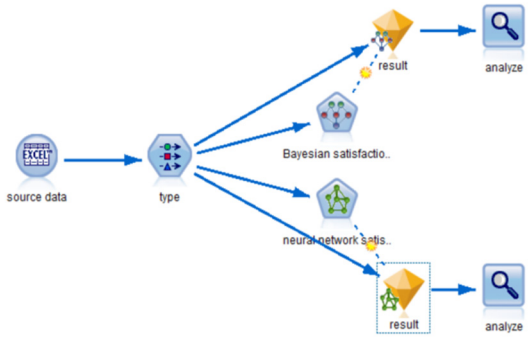

Figure 1. Modeling Flow Chart

3.2. Bayesian Network Model

Import the collated product review data into IBM SPSS Modeler software, select the data type, and predict satisfaction as the output variable. The other five variables are input variables, and the variable

type is set. Select TAN as the model structure type, and the parameter learning method is maximum likelihood. The constructed Bayesian network model is shown in Figure 2.

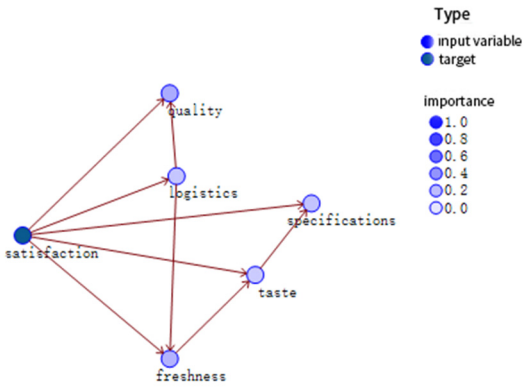

Figure 2. Bayesian Network Model Diagram

In the Bayesian network model, the steel-blue circle represents the target variable, the blue circle represents the input variable, and the depth of the blue circle color represents the importance of the input variable. The darker the color, the higher the importance. In the Bayesian network model diagram, satisfaction score as the target variable is connected to the other five variables through arrows. From the diagram, it can be seen that satisfaction as the target variable has relationships with the other five variables, quality is related to logistics, specifications are related to taste, taste is related to freshness, and freshness is related to logistics.

The conditional probability distribution diagram of the model output is shown in Figure 3.

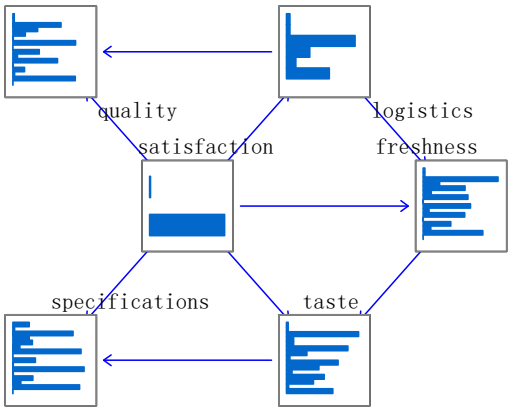

Figure 3. Bayesian Network Conditional Probability Distribution

The conditional probability distribution diagram clearly shows the interrelationships between various factors. Satisfaction, as the target variable, is the parent node of other factors. The arrows between factors indicate both the parent-child relationship between factors and the interrelationships between them. For example, satisfaction is the parent node of freshness, which in turn is the parent node of taste, while there is also a relationship between logistics and quality. The interrelationships between factors form the structural model of satisfaction.

According to the model, the overall conditional probability table is shown in Table 3.

It can be seen that the probability of a score of 5 for the satisfaction node is the highest, reaching 0.986. This indicates that consumers have a high level of overall satisfaction with this product.

Table 3. Table of Conditional Probability of Overall Satisfaction Node

Satisfaction	1	2	3	4	5
Probability	0.000	0.000	0.006	0.008	0.986

The following is the order of importance of each factor:

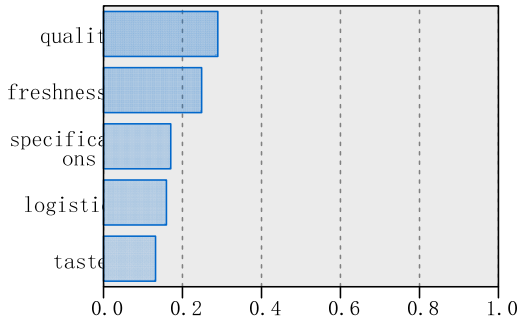

Figure 4. Importance Ranking of Satisfaction Factors Based on Bayesian Network

From the importance ranking diagram of factors affecting customer satisfaction with the product, it can be seen that the top three concerns of customers are quality, freshness, and specifications, followed by logistics and taste. It can be seen that the quality of this hairy crab has the greatest impact on customer satisfaction.

3.3. Neural Network Model

Using the neural network model for analysis, the satisfaction level is the output layer, and the various influencing factors are used as the input layer. The construction of the neural network is shown in Figure 5.

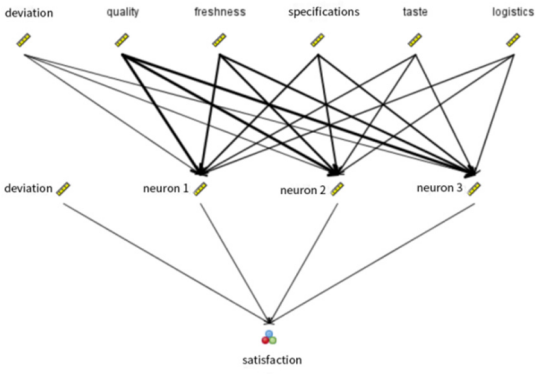

Figure 5. Neural Network Model

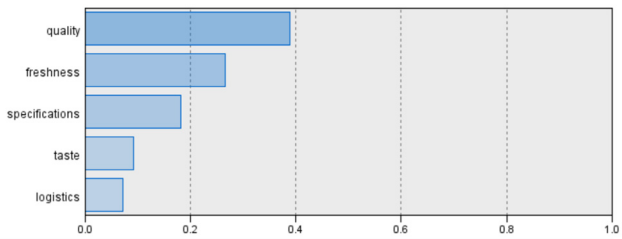

Figure 6. Importance Ranking of Satisfaction Factors Based on Neural Network

Using the neural network model, the factors that affect customer satisfaction were studied and the variable importance ranking shown in Figure 6 was obtained. From the figure, it can be concluded that customers are more concerned about product quality, freshness, and specifications, followed by factors such as taste and logistics.

4. Summary of Customer Satisfaction Model

For the customer satisfaction measurement models established by various models, the results they output are mostly consistent but there are also some differences. Therefore, we selected three indicators to compare the models. While ranking the importance of the influencing factors in the model output, each satisfaction model also outputs its own model accuracy rate, which is used as one of the model evaluation metrics. Then, using the test set, we predict the satisfaction level and obtain the accuracy and recall rates of each model. By comparing the three indicators, we obtain the optimal measurement model. The indicator comparison is shown in Table 4.

Table 4. Table of Conditional Probability of Overall Satisfaction Node

	Precision	Recall	Accuracy
Bayesian network model	0.861	0.743	0.892
neural network model	0.859	0.705	0.884

From Table 4, we can see that the accuracy of the Bayesian-based customer satisfaction model is 89.2%, the precision of the model during prediction is 86.1%, and the recall rate is 74.3%. In the neural network-based satisfaction model, the accuracy is 88.4%, the precision is 85.9%, and the recall rate is 70.5%. The accuracy of both models is above 85%, indicating that the ranking results of the factors affecting the model output are relatively reliable. Through the comparison of indicators, we can see that the indicator values of the Bayesian network customer satisfaction model are slightly higher than those of the neural network model, indicating that the Bayesian-based model is more superior.

The importance ranking of influencing factors obtained based on the Bayesian network satisfaction model is: quality, freshness, specifications, logistics, and taste; the importance ranking obtained based on the neural network satisfaction model is: quality, freshness, specifications, taste, and logistics.

5. Summary

Social progress has brought about a richness and demand for material life. People purchase goods through e-commerce platforms and other channels, resulting in a continuous growth of online review data generated from transactions. Customers express their true feelings through reviews. Based on online reviews from e-commerce websites, this article obtains factors that affect customer satisfaction with online fresh food purchases and their ranking through data mining. Based on the research results, several suggestions are proposed:

For customers, people who purchase goods will share their true usage experiences through reviews, providing other consumers with the most direct basis for making consumption decisions. By collecting relevant reviews on e-commerce platforms, customers can obtain factors that affect customer satisfaction and view evaluations from other customers to understand product quality, freshness, taste, and other aspects of performance. They can choose highly reputable and well-rated merchants for purchase. At the same time, attention should be paid to issues such as after-sales service and logistics to ensure their shopping experience and food safety.

For e-commerce platforms, in order to ensure that customers can purchase high-quality, fresh and delicious fresh products, they should strengthen the quality control of the merchants who have settled in, and ensure that the goods sold meet relevant national standards and regulations. Once any

merchants and goods that do not meet the requirements are found, they should be immediately removed from the market and handled accordingly. At the same time, e-commerce platforms should establish a sound freshness guarantee system to ensure that products can maintain freshness during transportation. In addition, taking fast and timely delivery measures is also a key link to ensure product freshness. Through efficient logistics services, customers can receive the fresh products they purchase in a timely manner, reducing waiting time and better enjoying the food. Finally, e-commerce platforms should encourage customers to evaluate and share the fresh products they purchase. These evaluations and sharing can help improve product transparency, provide feedback and suggestions for merchants, and promote their continuous improvement of product quality and services.

For manufacturers, firstly, they should strengthen product quality control, starting from the source, strictly screening and checking raw materials to ensure that the materials used are fresh, healthy and reliable. At the same time, they should strengthen the control of processing technology to ensure that the production process of the product complies with national standards and customer requirements, and take effective measures to maintain product freshness, such as using cold chain logistics and reasonably controlling the temperature and humidity of the storage environment. Secondly, they should set product specifications according to market demand and customer needs to avoid waste or shortage. Finally, manufacturers should optimize the taste of the product based on customer feedback and market demand. By collecting and analyzing customer evaluations and suggestions, they can understand consumers' preferences and needs for the taste of the product, and optimize it based on these feedbacks. Only in this way can they achieve greater success in the competitive market.

References

[1] Godes D, Mayzlin D. Using Online Conversations to Study Word of Mouth Communication, Marketing Science,2004, pp.545-560. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Bin F M H M, Aslinie B S M, Norashikin Y. Online commentaries of the sugar-sweetened beverages tax in Malaysia: Content analysis. Public health nursing ,Boston,2023.

[3] Tucci, L. A. Data mining techniques in CRM: Inside customer segmentation, K. Tsiptsis A. Chorianopoulos: book review. International Retail and Marketing Review. (2011).

[4] Archika S, Omair M S. A Comprehensive Artificial Intelligence Based User Intention Assessment Model from Online Reviews and Social Media. Applied Artificial Intelligence. (2022).