Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 4
ICCIA 2024

Research on influencing factors of pharmaceutical e-commerce sales based on web crawler and support vector machine

Shengbo Hu

Wuhan University of Technology, Wuhan, China

Abtract. Based on the current background of the rapid development of the pharmaceutical e-commerce industry, this paper provides an in-depth discussion of the factors affecting pharmaceutical e-commerce sales. With the help of Python crawler technology, the pharmaceutical e-commerce data of Alibaba Health platform is collected with GanMaoLingKeLi(999) as an example. Using support vector machine (SVM), according to the selected characteristic indicators, different schemes are set to predict the sales of pharmaceutical products and determine the influencing factors of pharmaceutical e-commerce sales. The results show that comment characteristics, transaction characteristics, shop characteristics, service characteristics and product characteristics all have different degrees of influence on product sales. Among them, the influence of comment characteristics is the most significant. Based on these results, merchants should attach great importance to the construction of review content and improve the quality of pharmaceutical services to attract and retain consumers. The government should promote the pharmaceutical e-commerce industry to improve service quality and encourage technological innovation. Third-party platforms should optimise website design and strengthen brand image through live broadcasting and other means.

Keywords: Web crawler, Support vector machine (SVM), Pharmaceutical e-commerce, Sales Analysis.

1. Introduciton

With the in-depth implementation of the “Action Programme for Promoting High-Quality Development of Health Industry” issued by the National Development and Reform Commission, the healthcare sector is witnessing a profound change. The Platform for Action explicitly proposes the establishment of a platform for cooperation between pharmaceutical distribution enterprises, medical institutions and e-commerce enterprises. It also strongly encourages the use of technologies like cloud computing and big data in drug distribution to enhance supply-demand information symmetry and transparency. Moreover, to boost medical efficiency and patient satisfaction, the policy also emphasises the acceleration of the development of pharmaceutical e-commerce, and encourages the provision of convenient services such as "online ordering (drug) shop pick-up" and "online ordering (drug) shop delivery"[1]. Driven by this policy, the "Internet + drug distribution" model represented by pharmaceutical e-commerce is accelerating, injecting new vitality into the innovation and development of the medical field. The 2019 COVID-19 outbreak spurred rapid growth in B2C pharmaceutical e-commerce platforms like Alibaba Health and JD Health. According to iiMedia Research's "2023-2024 Global and Chinese Pharmaceutical E-commerce Market and Development Trend Research Report", China's pharmaceutical e-commerce market reached 248.6 billion yuan in 2022, a year-on-year growth of 10%, and is expected to exceed 340 billion yuan in 2026[2]. At the same time, the pharmaceutical B2C market dominates e-commerce, exhibiting vast potential and rapid growth macroscopically. Microscopically, however, profitability remains a concern due to significant disparities among key players and other pertinent issues. Specifically, the average gross profit margin for pharmaceutical e-commerce stands at approximately 19.3%, with significant disparities observed among various business entities, ranging from a low of 6% to a high of 45%.[3].In this context, how do pharmaceutical e-commerce companies operate on representative e-commerce platforms? What factors influence their sales behavior? How significant is the impact of each influencing factor? To answer these questions and foster the healthy development of

pharmaceutical e-commerce, there is an urgent need to conduct a study on the factors that influence pharmaceutical e-commerce sales.

Currently, academic research on pharmaceutical e-commerce sales primarily centers on customer satisfaction and behavior, emphasizing the consumer perspective. However, there is limited exploration into the relationship between objective influencing factors and sales outcomes. Moreover, most empirical data in existing studies rely primarily on online reviews and user surveys, with insufficient attention to objective indicators like pharmaceutical consulting services, pharmacist qualifications, and medication guidance. Regarding research methodology, support vector machines are primarily utilized in e-commerce for demand forecasting and sentiment analysis, with limited application in studying sales influencing factors.

Given the preceding reasons, this paper focuses on GanMaoLingKeLi(999) sold on Alibaba Health, a leading B2C pharma e-commerce platform in China[3]. Using a Python web crawler, we collect the latest dataset, conduct thorough cleaning and preprocessing, and analyze sales influencing factors with Support Vector Machines in Machine Learning. Our aim is to promote optimal development in the pharmaceutical e-commerce industry. The marginal contributions of this paper include: 1. Innovating the research perspective by focusing on influencing factors of pharmaceutical e-commerce sales, overcoming the limitations of prior consumer-oriented studies (e.g., customer satisfaction and behavior). This objective analysis enhances understanding of the connection between influencing factors and sales, offering new theoretical support for industry development.Enhanced data collection and processing methods: Utilizing Python web crawler technology, this paper collects the latest dataset from Alibaba Health, a leading B2C pharmaceutical e-commerce platform in China, ensuring data timeliness and authenticity. Data cleaning and preprocessing effectively eliminate noise and outliers, improving analysis reliability. 3. Expanded research methodology: This paper applies the machine learning algorithm, support vector machine, to study pharmaceutical e-commerce sales influencing factors, broadening the algorithm's application scope in e-commerce. SVM's unique advantage in handling high-dimensional and nonlinear data effectively mines data relationships and reveals complex sales influencing factors.

2. Literature review

Driven by digitalization, pharmaceutical e-commerce, an emerging industry, is progressively becoming integral to the pharmaceutical field. Its rapid growth has led to increasingly profound academic research.

Regarding research content, some studies focused on the overview of pharmaceutical e-commerce's development, policies, challenges, and recommendations, affirming its potential while highlighting existing issues and proposed solutions. Dupuits[4] emphasized the significance of pharmaceutical e-commerce in healthcare, reducing costs and enhancing management efficiency. Yet, its growth has also given rise to challenges, including illicit drug trafficking and privacy violations. Orizio's[5] study revealed complex issues in online pharmacies, encompassing doctor-patient relations, consumer empowerment, drug quality, regulation, and public health impacts. Chen[6] noted that pharmaceutical e-commerce enhances drug distribution efficiency, yet faces challenges in China, including legal policy imperfections. Zhang[7], on the other hand, proposed a development path to build a pharmaceutical e-commerce ecosystem on the basis of these problems, which provides ideas for the healthy development of pharmaceutical e-commerce.

Concurrently, the proliferating number of pharmaceutical e-commerce users has sparked interest in user behavior studies. More literature delves into the purchasing behavior of these users, exploring sales influencing factors through the lens of customer satisfaction and consumer behavior. Among them, a significant focus is on the analysis of online reviews.. Under the big data environment, Huang et al [8] utilized ICTCLAS and AntConc to delve into online pharmacy review hotspots, studying their influencing factors, which aided online pharmacies in bolstering consumer trust and sales. Subsequently, Zhao et al [9] expanded this research, conducting a sentiment analysis of online

reviews, quantifying emotional tendencies and introducing five potential themes influencing customer satisfaction. Luo et al[10] discovered that the quantity and timing of additional reviews significantly impacted sales, utilizing Alihealth.tall.com data.

In addition, some of the literature also considered the service quality of pharmaceutical e-commerce.Yan et al.[11] analyzed the influence of service quality on customer satisfaction in pharmaceutical e-commerce using the SERVQUAL model across five dimensions, revealing the significance of perceived value. Wu et al.[12] built on this work, incorporating service encounter theory and the LSQ model to develop a comprehensive structural equation model exploring the intricate relationships between logistics service encounters, user perceptions, and customer satisfaction. Notably, Wu et al. found the path coefficient between logistics encounters and perceived risk to be insignificant, complementing Yan et al.'s insights on the role of perceived value. Zou et al[13] highlighted the need for pharmaceutical e-commerce companies to enhance medical consulting services to expand their market share through questionnaire surveys. He et al.[14] further examined this, drawing from literature, e-commerce characteristics, and the Chinese customer satisfaction index model. Their analysis revealed that service and perceived value directly and positively impact customer satisfaction, whereas brand image and quality assurance do not have a direct effect.

In terms of research methodology, the application of SVM in the field of e-commerce has been relatively mature. Based on the advantages of SVM's high prediction accuracy and robustness[15], SVM is often used for demand forecasting in e-commerce field. Chen et al.[16] proposed an LS-SVM model for e-commerce customer demand forecasting, leveraging the advantages of least squares support vector machine regression. On this basis,Tang[17] and Ma[18] applied SVM to dynamic demand prediction for fresh agricultural products, providing powerful tools for forecasting demand in the e-commerce context.

Meanwhile,SVM, as a classification model, is widely used in e-commerce sentiment analysis. Wang et al.[19] integrated Gabor and SVM for sentiment recognition, enhancing recommender system performance and user satisfaction. The study by Chen[20], on the other hand, proposed a DDAG-SVM-based online product review credibility classification model to classify the sentiment of user reviews on e-commerce platforms. This study not only inherits the results of previous researchers in the field of sentiment analysis, but also refines the sentiment analysis to the judgement of review credibility, which makes the study closer to practical applications. Zhang et al.[21] shifted their focus to quantifying behavioral characteristics of commenters and developed an independent SVM model to identify fake commenters. This study, although different from the previous sentiment analysis, is also an effective use of e-commerce reviews, reflecting the diversity and complementarity of research.

Overall, pharmaceutical e-commerce research spans macro and micro levels, including trends, policies, and consumer behavior. Micro-level studies often focus on consumer-driven sales factors, but comprehensive analyses of their objective impact on sales are scarce. Furthermore, existing research primarily relies on online reviews and user questionnaires, neglecting other objective influencing characteristics. As for research methodology, while SVM is widely used in e-commerce demand forecasting and sentiment analysis, its application in pharmaceutical e-commerce remains limited and underexplored.

3. Data and methods

3.1. Web crawler tool based on Python language

Web crawlers are programs that simulate human browsing to automatically gather web information. Python, with its numerous open-source libraries and frameworks, is a popular choice for such crawlers. These Python-based tools send HTTP requests to fetch HTML pages, extract required information, and convert it into the form required by the user for storage[22].As big data evolves and the "Three-Year Action Plan of 'Data Elements x' (2024-2026)" is issued by the National Data Bureau and other departments, emphasizing Data Elements x Commerce and Trade Circulation[23], data

elements gain increasing significance in business operations. Meanwhile, with the development of online shopping platforms, transaction volumes have skyrocketed. Using web crawlers to capture transaction data, analyze online transaction characteristics, and mine potential association rules can provide valuable insights for online shopping platforms and facilitate transactions for both parties. Currently, web crawler applications in e-commerce research are gaining popularity, with a focus on fresh food and cross-border e-commerce. However, there is a relative dearth of research on factors influencing sales in pharmaceutical e-commerce. Shi et al[24] obtained online review data based on web crawlers to compare different logistics service strategies in fresh food e-commerce supply chain and the differences between different fresh food products. Yang et al[25] used web crawler technology and text analysis to empirically examine the impact of listed companies' cross-border e-commerce business on corporate internationalisation based on 3.889 millions announcements issued by A-share listed companies from 2007 to 2020. Zhang et al[26] used web crawler technology and Word2vec model to establish a thesaurus of fresh food product features and construct a multivariate SVR demand forecasting model. Zhou[27] quantitatively evaluated the index of agricultural product circulation in each city using web crawler technology, and elaborated the marginal impact of rural e-commerce sinking on agricultural product circulation from the perspectives of differences in factor endowments, such as digital economy and public services, respectively.

3.2. Data collection and statistical analysis

1.The specific acquisition process of the dataset. Use Fiddler to get the html page of Mobile Taobao, use python library and framework to request the data through request, and match and filter according to the keywords. This dataset uses "GanMaoLingKeLi(999)" as the keyword to capture all real-time sales of GanMaoLingKeLi(999) products on Mobile Taobao. After filtering the keywords, the data is traversed to select the fields we need, including sales, price and other related information, as well as review information, and finally saved to excel.

2.Data Cleaning. The dataset was collected into three times, respectively, the sales price of the product, the detailed information of the product and the review information of the product. After three separate collections, the data of the three times are matched according to the product id. Since the collection has time interval and some products have been taken off the shelves, the products with incomplete information collection are eliminated, and some products with incorrect information are cleaned, such as the shop rating is 0 and the number of pictures is a decimal number. After the data collection is completed, the data is stored as a list. Subsequently, this list is saved to an Excel sheet using the pandas library, making it convenient to utilize the Excel sheet for data cleaning purposes. Finally, 1858 product information of GanMaoLingKeLi(999) is obtained.

3.Variable selection and data format conversion. On the basis of the completion of data cleaning, the required variables are selected for classification: (1) Comment characteristics. Including the number of comments (cumulative number of comments), the length of comments (the average of the number of words of all comments), the number of pictures; (2) Transaction characteristics. Including delivery costs (unconditional free shipping = 2, conditional free shipping = 1, no free shipping = 0), quality assurance (7 days no reason to return = 2, other guarantees = 1, does not support 7 days no reason to return = 0); (3) Shop characteristics. Including the degree of product conformity to description, logistics level, and service level; (4) Service characteristics. Including the presence of pharmacist consultation (yes = 1, no = 0), the presence of pharmacist qualifications (yes = 1, no = 0), the presence of medication guidance (yes = 1, no = 0); (5) product characteristics. Including product price, sales volume (with the number of payers as an indicator). Data format conversion is achieved through the filter and replace function of excel table.

4.Statistical analysis. Import the data into PyCharm. Utilize Python's data analysis library to conduct statistical analysis, distribution analysis, and correlation analysis on features such as comment characteristics, transaction characteristics, and other relevant characteristics in sequence.

3.3. Support Vector Machine (SVM)

Support vector machine (SVM) was first proposed by Cortes and Vapnik[28], which is mainly applied to small-sample, high-dimensional, nonlinear data classification problems[29]. Currently, support vector machines and random forests are the most accurate classification tools, with classification results significantly higher than many other classifiers[30]. However, there is still less research on applying support vector machines to the pharmaceutical e-commerce field.

SVM is a binary classification model, its basic model is defined as a linear classifier with maximum intervals on the feature space, and it also includes the kernel function trick, which makes it essentially a non-linear classifier.SVM encounters two situations when faced with a data classification problem, one in which the sample data is linearly separable, and the other in which the sample data is non-linearly separable. When the sample data is linearly separable, it will find an optimal classification boundary (hyperplane) that divides the data into two classes while maximising the geometric separation between the sample points and the classification boundary. When the sample data is non-linearly differentiable, the SVM will map the non-linear sample data to a higher dimensional space by using a kernel function, thus transforming the sample data into linearly differentiable, and then using the linearly differentiable method for classification.

As shown in Figure 1, the sample is linearly separable. Using the equation $wx + b = 0$ as the classification boundary, it divides the data into two categories: solid circles and hollow circles. Meanwhile, the points on $wx + b = 1$ and $wx + b = -1$ are support vectors, which are the sample points that play a crucial role in determining the location of the decision boundary of the SVM. However, in reality, there is almost no completely linearly separable data, in order to solve this problem, the concept of "soft spacing" is introduced, i.e., some points are allowed to fail to satisfy the constraints and are judged to be on the wrong side. The penalty factor C is used to control the tolerance of the model to misclassification. When the penalty factor is large, the SVM will try to ensure that all samples are correctly classified, even though this may lead to complex decision boundaries and easy overfitting. Conversely, when the penalty factor is small, the SVM will allow some samples to be misclassified in exchange for a simpler, more generalised decision boundary.

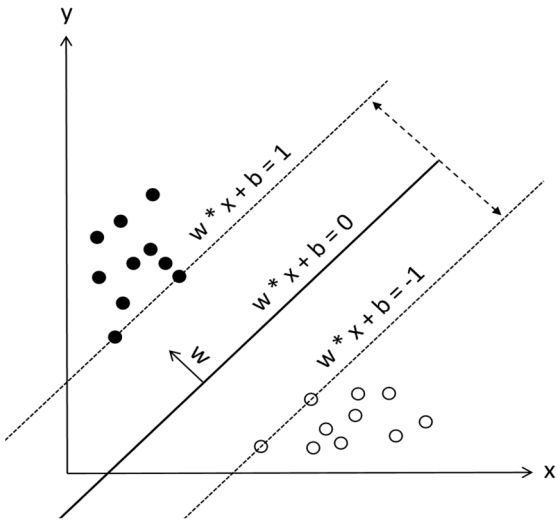

Fig. 1 Schematic diagram of support vector machine

3.4. Implementation Method of Classification Based on Support Vector Machine (SVM)

1.Determine the feature indicators. Select the indicators that may affect the sales of pharmaceutical products from the crawled and cleaned data set, and categorise them into comment characteristics, transaction characteristics, shop characteristics, service characteristics and product characteristics, and the specific indicators of each feature are shown in the table below:

Table 1. Variables and characteristic indicators affecting sales of pharmaceutical goods

influencing factors	Characteristic indicators
Comment characteristics	Number of comments, Length of comments, Number of pictures
Transaction characteristics	Delivery costs, Quality assurance
Shop characteristics	The degree of product conformity to description, Logistics level, Service level
Service characteristics	The presence of pharmacist consultation, The presence of pharmacist qualifications, The presence of medication guidance
Product characteristics	Product price

2.Classify the sales of goods as labels, while normalising the feature indicators. In order to simplify the study, it is assumed that the sales data displayed on Taobao is the real sales of pharmaceutical products, and there is no brushing and other behaviours. Calculate the median sales volume, and use the median as a classification criterion to classify the sales volume into 0 and 1. At the same time, the feature metrics are normalised, and all the features are adjusted to the same scale, so that each feature can be treated fairly during the model training. MinMax normalisation is performed using the MinMaxScaler function that comes with the python library, preserving the relative size relationship of the original data.

3. Determine the training and testing sets. After cleaning, there are a total of 1858 valid data. In order to improve the effectiveness of the model, the sample data was classified in a 1:1 ratio and divided into training and testing sets. Therefore, both the training and testing sets have 929 entries.

4. Determine the kernel function as well as the penalty factor. For linearly indivisible problems, the kernel function is needed to map the input data to a higher dimensional space, making it easier to partition the data linearly in the new feature space, and thus converting non-linearly divisible data to linearly divisible data. Whereas the penalty factor is used to weigh the loss and classification interval, a balance needs to be found between the complexity of the model and its generalisation ability. Therefore a grid search method is adopted to search the defined kernel function as well as the network of penalty factor parameters to find the factor with the highest model accuracy and output it.

5. Training model. Based on the training set, construct the SVM model for training, test the test set after training is completed, and output the model accuracy based on the test results.

6. Try different feature indicators and observe the accuracy rate. Delete one of the five feature indicators affecting sales in the training set in turn, retrain the model for testing, and observe the change in model accuracy.

4. Analysis of data characteristics

4.1. Statistical analysis of characteristic indicators

In terms of review characteristics, although the average number of reviews reaches 75.6, the standard deviation is as high as 158.1, suggesting that the number of reviews varies greatly between products. The mode number of 10 suggests that some shops may encourage users to post a specific number of reviews through some kind of strategy (e.g., incentives). The average comment length is moderate at 21.5 words, but the minimum value is only 4 words and the maximum value reaches 65 words, which indicates that there is a significant difference in the level of detail of users' comments. Meanwhile, the mean value of the number of images is 46.3, but the standard deviation reaches 99.9, indicating that the distribution of the number of images is extremely discrete and fluctuates greatly, which may be more related to the number of comments in different shops.

In terms of transaction characteristics, the average value of express delivery cost is 1.4, which is close to the value of "conditional free shipping", and most orders may need to fulfil certain conditions in order to be eligible for free shipping. However, the mode is 2, which means that the number of orders

with unconditional free shipping is also considerable, indicating that pharmaceutical products are in line with the majority of products on Taobao in terms of courier fees. The average value of quality assurance is only 0.1, which is much lower than the value of "7 days no reason to return", which is due to the special nature of pharmaceutical commodities. Pharmaceuticals belong to a special category, according to the "Code of Practice for the Quality Management of Pharmaceutical Business", in addition to the quality of the drug, once sold, no return.

In terms of shop characteristics, both product, service and logistics levels are relatively consistent, with a mean value of 4.8, while the standard deviation is 0.1, indicating that the differences between different shops are also relatively small.

In terms of service characteristics, the mean value of pharmacist counselling service is only 0.1, indicating that fewer shops provide this service. However, the mean value of pharmacist qualification reached 0.6, indicating that more than half of the shops had pharmacist qualification. This difference may reflect the different focus of shops in providing professional services. At the same time, about half of the shops provide medication guidance services, which is important for improving users' purchasing experience and ensuring medication safety.

In terms of product characteristics, the standard deviation of product prices is 34.4, while the maximum and minimum values differ significantly, which shows that the price difference between different products is more significant, which may be related to whether the products provide the above service characteristics and the content of the transaction characteristics.

Table 2. Descriptive statistics of GanMaoLingKeLi(999) commodity characteristic indexes

Influencing variable	Characteristic indicators	Average	Standard deviation	Minimum	Maximum	Median	Mode	Number of mode
Comment characteristics	Number of comments	75.6	158.1	1	2054	24	10	79
	Length of comments	21.5	8.7	4	65	19.8	16	49
	Number of pictures	46.3	99.9	0	900	12	0	308
Transaction characteristics	Delivery costs	1.4	0.9	0	2	2	2	1304
	Quality assurance	0.1	0.3	0	2	0	0	1642
Shop characteristics	The degree of product conformity to description	4.8	0.1	4.4	5	4.8	4.8	1382
	Logistics level	4.8	0.1	4.5	5	4.8	4.8	1500
	Service level	4.8	0.1	4.5	5	4.8	4.8	1509
Service characteristics	The presence of pharmacist consultation	0.1	0.3	0	1	0	0	1707
	The presence of pharmacist qualifications	0.6	0.5	0	1	1	1	1182
	The presence of medication guidance	0.5	0.5	0	1	0	0	936
Product characteristics	Product price	27.9	34.4	2	690	18.5	15	56

Note: delivery costs (unconditional free shipping = 2, conditional free shipping = 1, no free shipping = 0); quality assurance (7 days no reason to return = 2, other guarantees = 1, does not support 7 days no reason to return = 0);with or without pharmacist consultation (yes = 1, no = 0); the presence of pharmacist consultation (yes = 1, no = 0); the presence of pharmacist qualifications (yes = 1, no = 0); the presence of medication guidance (yes = 1, no = 0);

4.2. Distributional analysis of sales volume

In response to the distribution of sales of pharmaceutical goods, most of the sales of pharmaceutical goods (GanMaoLingKeLi(999)) are concentrated in the range of 0-10, occupying 71.04% of the number of goods. This indicates that the sales volume of this commodity is relatively low in most shops. While the number of items with sales ranging between 1,000-10,000 accounted for 2.74% of the total, the total sales volume reached 155,000, indicating that a very small number of shops have large sales volume, such as Alibaba Health Pharmacy, 999 Official Store and so on. Meanwhile the gap between shops is extremely large.

Table 3. Distribution of GanMaoLingKeLi(999) Sales

Volume range	quantities	Proportion	Average salesvolume	Total sales volume
0-10	1320	71.04%	1.57	2069
10-100	355	19.11%	33.24	11799
100-1000	132	7.10%	243.18	32100
1000-10000	51	2.74%	3039.22	155000

4.3. Correlation analysis between characteristic indicators and sales volume

All characteristics are positively correlated with sales, indicating that the selected characteristics are to some extent related to the increase in sales. The correlation coefficient between the number of reviews and sales is 0.354, indicating a more significant positive correlation between the number of reviews and sales. This means that the higher the number of reviews for an item, the higher its sales are likely to be. This may be because a high number of reviews reflects that the item receives a high level of attention or that consumers are more satisfied with the item. The correlation coefficient between the presence of pharmacist advice and sales is 0.242, showing a somewhat positive correlation. The presence of pharmacist advice may have increased consumer confidence in the product, which in turn contributed to the increase in sales. The correlation coefficient between the number of pictures and sales was 0.226, again showing a positive relationship. The number of images of the product may have influenced consumers' intuitive understanding of the product, and the more images there were, the more they may have attracted consumers' attention, which in turn boosted sales. The correlation coefficients of features such as length of comments and availability of medication instructions with sales are all less than 0.1, and their impact on sales is relatively small. This may be because these features are not a major factor in consumers' purchasing decisions, or because these features do not vary much between products and therefore have a limited impact on sales.

Table 4. Correlation between characteristic indicators and sales volume

Characteristic indicators	correlation coefficient	positive or negative correlation
Number of comments,	0.354	positive correlation
The presence of pharmacist consultation,	0.242	positive correlation
Number of pictures	0.226	positive correlation
Length of comments,	0.081	positive correlation
The presence of medication guidance	0.081	positive correlation
Product price	0.077	positive correlation
The presence of pharmacist qualifications,	0.066	positive correlation
Delivery costs	0.037	positive correlation
Service level	0.018	positive correlation
The degree of product conformity to description	0.011	positive correlation
Logistics level	0.008	positive correlation
Quality assurance	0.008	positive correlation

5. Empirical programme and results

According to the above SVM implementation method, the train_test_split function is used to divide the training set and the test set into 1:1, and at the same time, the random seed is set to be 42 to ensure

the repeatability of the experiment. At the same time, define the parameter grid to be searched, including the range of the penalty factor C, the type of the kernel function kernel and the size of the gamma parameter in the kernel function, through the grid search, and finally get the optimal parameter combination of the penalty factor C = 100, the kernel function kernel = 'rbf', the kernel function parameter gamma='scale'. According to the best combination of parameters to train the model, to get the accuracy of the model containing all the feature values, and then sequentially delete one of the five feature indicators in the practice set affecting the sales volume, retrain the model to test the accuracy of the model, the final results are shown in the table below:

Table 5. SVM classification accuracy for different schemes

Schemes	Situation of influencing factors	Modelling accuracy
Scheme 1	All factors	70.94%
Scheme 2	Delete comment characteristic	64.48%
Scheme 3	Delete transaction characteristic	69.21%
Scheme 4	Delete shop characteristic	69.43%
Scheme 5	Delete service characteristic	67.17%
Scheme 6	Delete product characteristic	70.72%

From the above empirical results, the following five conclusions can be analysed to determine the influencing factors of pharmaceutical e-commerce sales and their importance.

1. when the training set of SVM model includes all the influencing factors, the accuracy rate of model classification is the highest. When a certain influencing factor is removed, the accuracy rate of the model all decreases to a certain extent, thus indicating that review features, transaction features, shop features, service features, and product features all influence sales to a certain extent.

2. When the review feature is deleted from the training set, the accuracy rate of the model is significantly lower than that of the other influencing factors, which indicates that the review feature has the most significant effect on the sales of pharmaceutical products compared with the other four features.

3. When product features are deleted from the training set, the accuracy rate of the model is significantly higher than that of the other influencing factors, which suggests, to some extent, that the other product features have a weaker impact on the sales of pharmaceutical products than the other four features.

4. When transaction features, shop features or service features are deleted from the training set, the accuracy rate of the model is between that of deleting review features and deleting product features, which indicates that the degree of influence of transaction features, shop features and service features on the sales of pharmaceutical products is between that of review features and product features.

5. When the training set includes all the influencing factors, the accuracy rate of the model is 70.94%, and there is still 30% room for improvement, which indicates to a certain extent that there are other factors influencing the sales of pharmaceutical commodities in addition to the five types of factors involved.

6. Conclusion

The empirical findings reveal that multiple factors, including online shop reviews, transaction logistics and quality assurance, shop qualifications, pharmaceutical service quality, and product pricing, influence consumer purchasing behavior in pharmaceutical e-commerce. Notably, online reviews exert the most significant impact. Consumers often encounter information asymmetry when purchasing pharmaceuticals, so they often rely on online reviews for decision-making. Online reviews provide direct feedback on products, services, and shops, reducing purchase risks and enhancing understanding of other buyers' experiences. Addtionally, the quantity and content of reviews reflect consumer attention and recognition, with a high volume indicating widespread willingness to share and enhancing review credibility. Positive comments generally enhance trust in products and shops.

Simultaneously, the quality of medicine-related services also impacts consumer satisfaction, which in turn influences consumer purchasing behaviour. The correlation coefficient between the indicator of service characteristics, with or without pharmacist consultation feature, and sales volume is second only to the number of reviews. This highlights the significance of medical consulting services in consumers' purchase of pharmaceutical goods. Given the close relationship between pharmaceutical products and consumer health, consumers exhibit caution in purchasing, demanding not only product information but also professional guidance. Pharmacists, as authorities in the field, provide accurate information on medication use, mitigating risks. The availability of pharmacist consultation services significantly impacts consumer purchases, addressing their need for high-quality medicines and professional advice, which can encourage consumers to purchase pharmaceutical products with greater confidence.

Finally, the sales of pharmaceutical products are influenced by logistics, quality assurance, operational qualification, and price, but these factors have limited impact. With the maturity of e-commerce, logistics and quality assurance have become basic standards, and consumers prioritize differentiated services like medical consultation. Additionally, the studied dataset focuses on a common OTC medication, GanMaoLingKeLi(999), minimizing the impact of logistics and quality on purchasing decisions. While operational qualification is crucial for assessing merchant reliability, consumers may prioritize platform reputation and product quality. Finally, while price is a consideration, consumers often prioritize medicine quality and effectiveness, limiting the impact of price on sales.

7. Recommendations for the key players involved in pharmaceutical e-commerce

For merchants selling on platforms, it is essential to enhance their sales strategies amidst the information overload consumers face. Reviews are vital for consumers to assess products and merchants to showcase service quality. Encouraging user feedback, no matter positive or negative, and responding promptly, which can improve services and reduce information asymmetry. As a result, it reduces the risk of users purchasing goods and promotes the generation of user purchasing behaviors. Additionally, merchants should enhance pharmaceutical support services, offering personalized online consultation and employing more licensed pharmacists[31]. These efforts enhance the shopping experience, build trust, and ultimately drive sales.

As a policy maker, the government should promote service quality improvement in the pharmaceutical e-commerce industry. The government should actively promote the cooperation and exchange between the pharmaceutical e-commerce industry and other industries. It can establish cooperative relationships with medical institutions and insurance companies to create a one-stop medical and healthcare service platform. This can be achieved by facilitating the circulation of personal health data files and other digital elements[23]. The aim is to provide consumers with more comprehensive and convenient medical and healthcare services. Secondly, the government can encourage and support pharmaceutical e-commerce platforms to innovate in the government should encourage technological innovation in pharmaceutical e-commerce, integrating big data and artificial intelligence to construct patient profiles and inform health and medication decisions based on purchase records and usage cycles.

For third-party platforms, improving website design and brand image is crucial to enhance user experience and competitiveness. It is recommended that information such as reviews and medical consulting services be given more prominence on the platform when designing the website.This will allow consumers to quickly access the information they need. Simultaneously, ensuring user-friendliness and aesthetics is important. Additionally, implementing a drug traceability system is crucial to uphold standards, prevent expired, counterfeit, or substandard drugs, and enhance brand image. Simultaneously, the platform can utilise live broadcasting to increase its visibility. Although e-commerce live broadcasting is prevalent in the retail industry, its application in the medical field is still in its infancy[32]. The platform can invite professional pharmacists or doctors to provide detailed

explanations of medicines, including their efficacy, applicable population, etc., so that viewers can gain an understanding of the professional knowledge of medicines.

References

[1] National Development and Reform Commission. Action Plan for Promoting High-Quality Development of Health Industry (2019-2022): Development and Reform Society No.1427 [EB/OL]. (2019-9-30) [2024-4-1].https://www.gov.cn/xinwen/2019-09/30/content_5435160.htm

[2] iiMedia Consulting. 2023-2024 Global and Chinese pharmaceutical e-commerce market and development trend research report [DB/OL]. (2023-10-12) [2024-4-1].https://www.iimedia.cn/c400/96151.html

[3] Zhou Yutao. ' Fourth Terminal ' led the industry 2022 ~ 2023 China pharmacy pharmaceutical e-commerce development report [J]. China pharmacy, 2023, (04) : 103-111.555

[4] Dupuits, FMHM.The Effects of the Internet on Pharmaceutical Consumers and Providers[J].Disease-Manage-Health-Outcomes,2002,10(11),679–691.

[5] Orizio,G.,et al.Cyber drugs:A Cross-Sectional Study of Online Pharmacies Characteristics[J].The European Journal of Public Health,2009,19(4),375–377.

[6] Chen Debao. Problems and improvement measures in the development of pharmaceutical e-commerce in China [J].Foreign economic and trade practice, 2016, (01) : 38-40.

[7] .Zhang Xiaheng.Construction and development path of pharmaceutical e-commerce ecosystem under ' Internet + ' [J].Contemporary Economic Management, 2016,38 (11) : 26-29.DOI : 10.13253 / j.cnki.ddjjgl.2016.11.005.

[8] Huang Zhe, Li Hui. Influencing factors of online reviews of online pharmacies based on big data [J].Journal of Shenyang Pharmaceutical University, 2016,33 (10) : 833-838.DOI : 10.14066 / j.cnki.cn21-1349 / r.2016.10.013

[9] .Zhao, Xiangqi; Gao, Lixiang; Huang, Zhe.Customer satisfaction evaluation for drugs: A research based on online reviews and PROMETHEE-II method[J].Plos One,2023,18(6):0283340.

[10] Luo, Yumei; Li, Yuwei; Ye, Qiongwei.Impacts of Online Additional Reviews on the Sales Volume of Cross-Border Pharmaceutical E-Commerce Platforms[J].Journal Of Global Information Technology Management,2022,25(1),83-101.

[11] Yan, Ma Jing; KANG, Taewon.A Study on the Effects of Service Quality of Pharmaceutical E-commerce on Customer Satisfaction -Mediating Effect of Perceived Value-[J].Korean-Chinese Social Science Studies,2021,19(3),185-205.

[12] Wu, Jianyun; Dong, Mingqiu.Research on customer satisfaction of pharmaceutical e-commerce logistics service under service encounter theory[J].Electronic Commerce Research And Applications,2023,58:101246.

[13] Zou Yueqing, Fan Mengyuan, Xu Xiaoyi, et al. ' Internet + ' under the background of medical e-commerce consumer group characteristics positioning and development strategy research-based on a sample survey of permanent residents in Wuhan [J].Modern Business, 2018, (25) : 26-28.DOI : 10.14097 / j.cnki.5392 / 2018.25.007.

[14] He Yufang, Ma Xinyu, Cui Yanyin, et al. Research on customer satisfaction of pharmaceutical e-commerce platform based on Chinese customer satisfaction index model [J]. China Pharmaceutical, 2023,32 (04) : 1-6.

[15] Jiang Feng, Zhang Wenya. Application of machine learning methods in economic research [J]. Statistics and Decision Making, 2022,38 (04): 43-49.DOI: 10.13546 / j.cnki.tjyjc.2022.04.008.

[16] Chen, Qisong; Wu, Yun; Chen, Xiaowei.Research on Customers Demand Forecasting for E-business Web Site Based on LS-SVM[J].Proceedings Of The International Symposium On Electronic Commerce And Security,2008,66-70.

[17] Tang Yifei. Dynamic demand forecasting of fresh agricultural products in e-commerce environment [D].Nanjing University, 2014.

[18] Ma Jiayu. Research on the prediction of fresh agricultural products demand of a farm e-commerce based on combination model [D]. Shandong University of Science and Technology, 2020.DOI : 10.27275 / d.cnki.gsdku.2020.000887.

[19] Wang Gang; Yin Fenxia.A Method of Recommendation Based on Affection Semantic in E-Business[J].Software Engineering And Knowledge Engineering: Theory And Practice, Vol 1,2012,114,369.

[20] Chen Yanfang. Online product review credibility classification model based on DDAG-SVM [J].Information theory and practice, 2017,40 (07) : 132-137.DOI : 10.16353 / j.cnki.1000-7490.2017.07.024.

[21] Zhang Wenyu, Yue Kun, Zhang Binbin. Detection of fake reviewers in e-commerce based on D-S evidence theory [J].Microcomputer system, 2018,39 (11) : 2428-2435.

[22] Sasi A, Deep A, Kumar K, et al. Machine Intelligence and Smart Systems[C].Singapore.Springer,2021:287-296.https://doi.org/10.1007/978-981-33-4893-6_26

[23] National Bureau of Data. " Data Elements × 3-Year Action Plan (2024-2026) " : Country Number Policy No.11 [EB/OL]. (2024-1-5) [2024-3-23].https://www.cac.gov.cn/2024-01/05/c_1706119078060945.htm

[24] Shi Chengyu, Chen Guaiheng, Wang Yan, etc. Research on logistics service strategy of fresh e-commerce supply chain from the perspective of big data [J].Agricultural technology and economy, 2023, (10) : 129-144.DOI : 10.13246 / j.cnki.jae.2023.10.005.

[25] Yang Shenggang, Xie Jinyuan, Cheng Cheng. Cross-border e-commerce, supply chain optimization and enterprise internationalization - empirical evidence based on big data text analysis [J]. International trade issues, 2023, (10) : 1-18.DOI : 10.13510 / j.cnki.jit.2023.10.001.

[26] Zhang Yanliang, Dai Peipei. Multivariate SVR demand forecasting for fresh products-Extraction of customer perception factors based on online reviews [J].Journal of China Agricultural University, 2022,27 (07) : 275-282.

[27] Zhou Wei. The impact of e-commerce sinking on the circulation of agricultural products from the perspective of factor endowment differences [J].Business Economics Research, 2022, (01) : 89-92.

[28] Cortes, C., Vapnik, V. Support-vector network [J]. Machine Learning, 1995(20) : 1-25.

[29] Gu Shen, Wang Shujuan. Research on carbon financial risk early warning model based on SVM [J].East China Economic Management, 2019,33 (03) : 179-184.DOI : 10.19629 / j.cnki.34-1014 / f.171220004.

[30] Cernadas E., Amorim D. Do we need hundreds of classifiers to solve real world classification problems? [J]. Journal of Machine Learning Research, 2014(1) : 3133-3181.

[31] Zhu Wenjing, Liang Miao. Based on the 5G era, thinking about the pharmaceutical service mode of online pharmacies in China [J].Exploration of rational drug use in China, 2021,18 (06) : 10-14.

[32] Zhou Yutao.B2C to the left, O2O to the right 2021 Report on the development of pharmaceutical e-commerce in Chinese pharmacies [J].Chinese pharmacies, 2022 (04) : 94-99.