International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 3556-3565

https://doi.org/10.70517/ijhsa463300

Analysis and Prediction of E-commerce Users' Purchasing Behavior Based on Logistic Regression Models

Wenrui Xu1,*

[1] Guangdong Polytechnic of Science and Technology, Dongguan, Guangdong, 523000, China

Corresponding authors: (e-mail: xuwrr001@163.com).

Abstract In this paper, we use the real data provided by Tianchi Big Data Research Platform to predict which commodities will be purchased by this user in the short term among the commodities that the user has interacted with. Firstly, the collected historical interaction data of commodities are normalized by Z-score. Then the features are extracted for coding, and the number of features is reduced based on the chi-square test method to improve the modeling efficiency and accuracy. Finally, the processed user purchase behavior data is input into the logistic regression model for e-commerce user purchase behavior prediction. The AUC value of the logistic regression model is greater than 0.5, the percentage of the number of purchasers increases with the increase of the predicted probability value, and the number of non-purchasers decreases with the increase of the probability value of the score segment. The model prediction results are consistent with the actual purchase, and the model is valid.

Index Terms Z-score, chi-square test, logistic regression, user purchase behavior prediction

I. Introduction

Accompanied by the arrival of the information age in the 21st century, the development of network information technology and platforms, people's quality of life and living standards have improved dramatically, and subsequently brought about a revolution in the way of life [1]. In such an era, e-commerce, a new modern business model came into being, which transforms the traditional consumption concept and transaction mode, and the economic structure then undergoes important changes [2], [3]. This is on the one hand to promote the development of the Internet economy, on the other hand, compared with physical shopping to save resources, optimize the purchase method, and help to protect the environment on which we depend [4]. As of December 2021, China's online shopping user scale reached 842 million, an increase of 59.68 million compared to December 2020, accounting for 81.6% of the overall Internet users [5]. However, as the number of e-commerce continues to increase, the variety of e-commerce platform commodities is getting richer and richer, which effectively meets the needs of people's product consumption, and the emergence of this situation is likely to increase the user's burnout of commodity purchase, which in turn reduces the turnover rate of commodities [6]. At this stage, in order to better meet the needs of the development of e-commerce, based on algorithms, the analysis and prediction of e-commerce user's purchasing behavior has become an extremely necessary initiative.

The purchasing behavior of e-commerce consumers is mainly characterized by personalization, bias towards service demand, initiative and rationalization [7]. In view of these characteristics, the literature [8] divided e-commerce consumers into four different online consumer groups, and then tested the conceptual model of the factors influencing the shopping behavior of each group separately. Literature [9] investigated the key factors influencing consumer sentiment and purchase behavior in e-commerce, including price, brand loyalty, product reviews, product quality, and promotions. Literature [10] found through a meta-analysis that survey trust, perceived risk, perceived security, and electronic word-of-mouth (e-WOM) were the key factors influencing consumers' decision-making on purchasing behaviors in e-commerce environments.

However, in recent years, with the gradual expansion of buyer-users on m-commerce platforms, the number of sizes of users' historical behavioral data has also increased dramatically. Problems such as data sparsity and low accuracy of collaborative filtering algorithms have become increasingly prominent, resulting in the inability to accurately predict user behavior [11]. Therefore, many researchers have proposed individual learning prediction models such as logistic (logistic) regression, support vector machine (SVM), multilayer perceptron and neural network [12]. Literature [13] analyzed and predicted the purchase behavior of Czech luxury goods using logistic regression analysis and found that older consumers were more concerned with brand loyalty, while younger buyers prioritized sustainability. Literature [14] constructed a prediction algorithm with optimized parameters for the purchasing behavior of electricity-using merchants by using a Support Vector Machine (SVM)-based model, which

further improved the model's prediction performance and overall prediction accuracy. Literature [15] used Gaussian Mixture Model (GMM) and Multi-Layer Perceptron (MLP) used to cluster customers and predict consumer behavior, and the results showed that the method achieved more than 95% accuracy. However, a single model has weak explanatory power and low accuracy in the prediction of user purchase behavior, and many scholars have found that the combination of models is better than a single model through research [16]. Literature [17] fused two heterogeneous algorithms of Logistic Regression and Support Vector Machine (SVM) to construct a user purchase prediction behavior model, which was used to predict whether users would produce purchase behavior for specific goods in the coming day, and the experimental results showed that the fusion model had better prediction effect compared with a single model. Literature [18] uses SelectKBest to improve the logistic regression model, after validation and comparative analysis, the fusion algorithm has different magnitudes of improvement in F1 value, precision and recall before and after feature selection. For the user purchase behavior prediction research, it can be seen that the single Logistic regression model and the combination of Logistic model prediction, have made to a certain progress, but still need to further in-depth research.

The study selects some real user-item behavior data of Alibaba Group from November 25 to December 3, 2024, and applies logistic regression to predict the list of user-items that appear to be purchased in the following week. The steps of the prediction model are divided into: first preprocess the numerical feature data using Z-score normalization. Then the category-based feature data are coded using 0-1 coding and chi-square test is performed. Subsequently balancing positive and negative samples and selecting samples for training and testing the model. Finally, the data were entered in the constructed logistic regression model for prediction.

II. Method

With the development of e-commerce, online shopping has become a major mode of consumption, compared with offline, online consumption has the advantages of low price, wide variety, convenient price comparison, and less influenced by business hours and geography [19]. In order to quickly and effectively locate the corresponding consumers from a sea of people, this paper adopts the real data provided by the Tianchi big data research platform, and uses the logistic regression model to predict which products will be purchased by users in the short term.

II. A. Data source selection and processing

II. A. 1) Introduction to data sources

Mobile turnover in Alibaba Group's 2024 Double 11 promotion exceeded 144.18 billion yuan. Compared with the PC era, the access to the mobile network is anytime and anywhere, with richer scenario data, such as the location information of the user, the time pattern of the user's access and so on. In view of this, Tianchi Platform Data Lab provides a part of relevant data: some real user-commodity behavior data of Alibaba mobile e-commerce platform from November 25 to December 3, 2024, based on which it predicts the list of user-commodities that appear to be purchased in the following week, in order to provide better services for mobile users. In order to better test the effectiveness of data mining, Tianchi provides online score query. The raw data has a total of 22256563 records, consisting of the purchase records of 10000 users obtained from random sampling. The number of product brands interacted is 296,506, and the number of product categories involved is 8633.

II. A. 2) Data processing

In the actual business raw data may have the following kinds of problems, null values, outliers, duplicates, etc. Data preprocessing is to organize and clean these data to make the data more standardized.

For linear models, feature normalization is to make different features in the same magnitude range, to avoid the data of very large magnitude and other constant magnitude data to cause misinterpretation level impact. If the original data is used for analysis, it will increase the role of indicators with large values in the analysis and weaken the role of indicators with small values in the analysis. Therefore, standardization can reduce the impact of different magnitudes on the results. And after standardizing the variables with different scales, it can reduce the influence of features with large variance and make the model more accurate. And it can speed up the convergence speed of the learning algorithm.

Z-score standardization is a common method of data processing, also called standard deviation standardization method, after Z-score standardization, the data will meet the standard normal distribution. The calculation formula is shown in equation (1):

$$Z_i = \frac{x_i - \bar{x}}{s} \tag{1}$$

where $x_i(i=1,2,3,\ldots,n)$ denotes the observation in the original data, $\bar{x}$ denotes the mean of this feature, and s denotes the standard deviation. z-score normalization method can enhance the accuracy of the model to eliminate the effect of magnitude and scale.

II. B. Category type feature processing and testing

II. B. 1) Category type feature processing

In the data used in this paper, the raw data of category type features are in the form of strings, which cannot be input into the model for training, and need to be encoded by encoding operation to convert them into numerical data. In this paper, 0-1 encoding is used for encoding.

II. B. 2) Chi-square test for category type characteristics

For regression and categorization problems features can be tested using, for example, chi-square tests.

The chi-square test is a commonly used hypothesis testing method to test whether there is a correlation between two categorical variables. The principle is to determine whether there is a significant relationship between two variables by comparing the degree of deviation between the actual observed values and the theoretical expected values [20].

The principle of chi-square test can be summarized in the following steps.

(1) Determine the research object: select the two categorical variables to be analyzed and collect the relevant observational data.

(2) Constructing hypotheses: establish the null hypothesis and alternative hypotheses, in which the null hypothesis refers to the absence of correlation between the two variables, and the alternative hypothesis is the opposite.

(3) Calculate the expected value: calculate the theoretical expected value, i.e., the joint distribution between the two variables if the null hypothesis is valid, based on the observed data.

(4) Calculate the chi-square value: Compare the actual observed value with the theoretical expected value to calculate the chi-square value, whose formula is shown in equation (2):

$$x^2 = \sum_{i=1}^{k} \frac{(A_i - E_i)^2}{E_i} \tag{2}$$

where $A_i(i=1,2,3,\cdots,k)$ represents the actual observed value of a category, $E_i(i=1,2,3,\cdots,k)$ represents the theoretical expectation of the i level, and k is the number of categories.

(5) Judging the significance: according to the relationship between the chi-square value and the degrees of freedom, the P-value of the chi-square distribution is calculated, and if the P-value is less than the significance level, the null hypothesis is rejected and the correlation between the two variables is considered to exist.

The application scenarios of the chi-square test include classification problems and feature selection problems. In classification problems, the chi-square test can be used to assess the correlation between each feature and the target variable, and select the most relevant feature to model the target variable. In feature selection problems, the chi-square test can be used to select the features that are most relevant to the target variable, thereby reducing the number of features and improving modeling efficiency and accuracy.

The chi-square test can be used to measure the correlation between a categorical variable and an outcome variable. Specifically, it can be used to calculate the chi-square value between each feature and the outcome variable, and then determine whether the feature is important or not based on the magnitude of the chi-square value for feature selection. That is, it can measure the degree of correlation between a feature and a labeled variable. Its application scenarios include the following.

(1) Feature selection: the chi-square test can be used to select features that have a significant effect on the target variable. When the feature space is very large, the chi-square test can be used to select the most relevant features, thus reducing computational costs and improving model performance.

(2) Hypothesis testing: The chi-square test can be used to verify certain hypotheses, such as determining whether two random variables are correlated.

(3) Data compression: The chi-square test can be used for data compression, which can be used to reduce the number of input features by selecting the features that are most correlated with the output variables. It should be noted that the chi-square test can only deal with categorical variables, for continuous variables need to be discretized.

II. C.Forecasting methodology

Logistic regression is a generalized linear model based on the fact that the dependent variable is binomially distributed Logistic regression is a very useful tool when there is a need to predict binary outcomes from a series of continuous or categorical predictor variables [21]. Logistic regression can easily solve categorization problems simply by using a logistic function on top of the general linear regression model. The logistic function is also known as the Sigmoid function and has the functional form:

$$g(z)=\frac{1}{1+e^{-z}} \tag{3}$$

In the collected data, each component represents a feature, and the purpose of logistic regression is to learn a 0/1 classification model from the features which is a linear combination of the features as an independent variable, since the independent variable takes on a range of values of $[-\infty,+\infty]$. Therefore, a logistic function is used to map the independent variables onto (0, 1) and the mapped values are considered to be the probability of belonging to $y=1$.

The linear combination of the predictor variables is as follows:

$$\theta_0+\theta_1 x_1+,...,+\theta_n x_n=\sum_{i=1}^{n}\theta_i x_i=\theta^T x \tag{4}$$

Step 1: Construct the prediction function for logistic regression:

$$h_\theta(x)=g(\theta^T x)=\frac{1}{1+e^{-\theta^T x}} \tag{5}$$

The value of $h_\theta(x)$ denotes the probability of $y=1$ for the input value x and the precondition that the parameter θ is predicted when the input predictor variables are fixed:

$$p(y=1\mid x;\theta)=h_\theta(x) \tag{6}$$

$$p(y=0\mid x;\theta)=1-h_\theta(x) \tag{7}$$

Step 2: Construct the loss function.

The probability of correct prediction for a single sample is:

$$p(y\mid x;\theta)=\left(h_\theta(x)\right)^y(1-h_\theta(x))^{1-y} \tag{8}$$

The probability distribution of the entire sample space is:

$$L(\theta)=\prod_{i=1}^{m}P(y_i\mid x_i;\theta)=\prod_{i=1}^{m}(h_\theta(x_i))^{y_i}(1-h_\theta(x_i))^{1-y_i} \tag{9}$$

To construct a convex loss function, taking the logarithmic expansion yields:

$$l(\theta)=\log L(\theta)=\sum_{i=1}^{m}(y_i\log h_\theta(x_i)+(1-y_i)\log(1-h_\theta(x_i)) \tag{10}$$

The final loss function is:

$$J(\theta)=-\frac{1}{m}[\sum_{i=1}^{m}(y_i\log h_\theta(x_i)+(1-y_i)\log(1-h_\theta(x_i))] \tag{11}$$

Step 3: Use gradient descent to find the minimum value.

It is necessary to find a θ^T such that $J(\theta)$ is minimized, and the formula for parameter iteration using gradient descent is as follows:

$$\theta_{j+1}=\theta_j-\alpha\frac{\partial}{\partial\theta_j}J(\theta)=\theta_j-\alpha\frac{1}{m}\sum_{i}^{m}(h_\theta(x^{(i)})-y^{(i)})x_j^{(i)} \tag{12}$$

III. Results and Discussion

The data processing methods of numerical and categorical features and the feature testing method based on chi-square test were introduced above respectively. This chapter first analyzes the data provided by the Tianchi platform, then builds a logistic regression model, and finally inputs the processed data into the model to predict user purchases.

III. A. Data analysis

III. A. 1) Visual analysis of user buying behavior

In order to better understand the data, this paper counts the four behaviors of consumers in 9 days: clicking, adding to cart, collecting and purchasing, and the trend of user behavior is shown in Figure 1. Among them, the number of clicks far exceeded the number of add-to-carts, favorites and purchases, and the behavior trend was stable from November 25 to the end of November, due to the fact that consumers had just experienced "Double 11" in November. The upward trend in early December was due to the fact that users were preparing for Singles' Day.

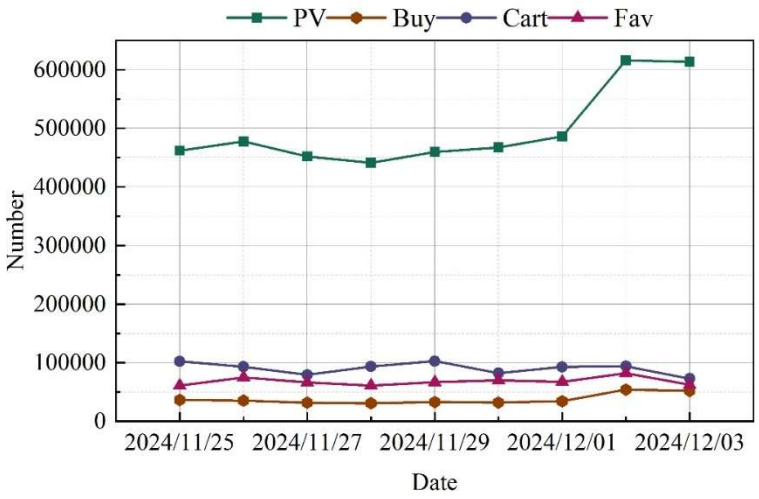

Figure 1: User behavior trends

Since there is no correlation between the two behavior types of collection and adding to cart, the user behavior paths are divided into clicking and then collecting before buying (path 1) and clicking and then adding to cart before buying (path 2). The user behavior conversion rates of the two paths are shown in Figure 2. User behavior path 1 purchase conversion rate of 8.3%, path 2 purchase conversion rate of 22.9%. This is because Taobao point to open the home page that has a shopping cart option, which is convenient for consumers to shop before the goods for further screening, and view favorites than view the shopping cart more than one step of the operation process, but also does not facilitate the user to a variety of goods for one-step purchase.

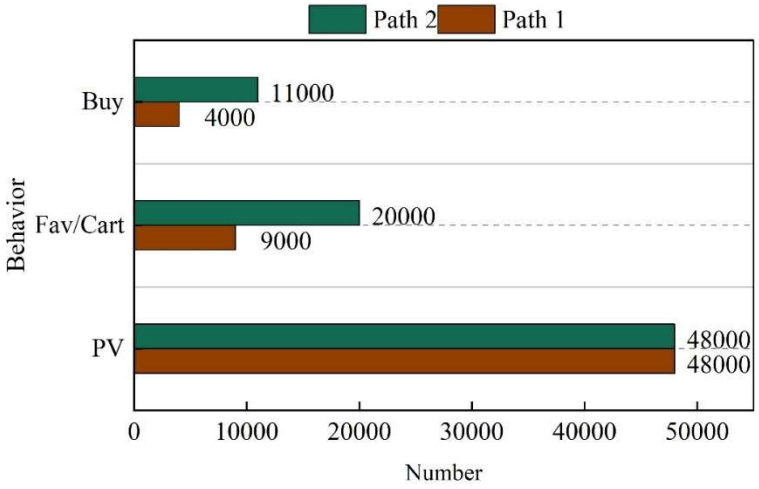

Figure 2: Rate of conversion of user behavior

III. A. 2) Constructing derived features

Consumer purchase behavior prediction belongs to the binary classification problem, construct the classification label, set the purchase behavior as 1, and the other three behaviors as 0. Firstly, construct the consumer characteristic group, commodity characteristic group and time characteristic group, among which the consumer characteristic group mainly constructs the number of clicks, add-ons, favorites, and purchases of consumers for specific commodities. Consumer conversion rate mainly includes add purchase conversion rate, collection conversion rate, purchase conversion rate. The behaviors are assigned different weight values, based on Figure 2 this paper assigns 1 point to click, 2 points to collect, and 3 points to add purchase, and weights all the behaviors of consumers corresponding to specific commodities to get a total score derived features. The commodity feature group mainly includes the number of times the commodity has been clicked, added to the shopping list, favorite, and purchased. The number of consumers who clicked, added, favorited, and purchased. Commodity heat, which is the proportion of sales of the commodity to the sales of the commodity category. The time feature cluster includes the date, day of the week, and hour features.

III. A. 3) Feature selection

The chi-square test is an important part of data analysis, through the chi-square test of the features, the features with higher chi-square values are selected as the features of the model for predicting the user's purchasing behavior, in order to reduce the burden of redundant features on the training of the model and to save the training time. The chi-square test results of the category type features are shown in Table 1. The degree of freedom is 1. According to formula (2), the chi-square values of click, add to cart, collection, consumer feature group, product feature group and time feature group are calculated, in which the chi-square of add to cart is the highest (23012.32).The six category-type features have asymptotic significance, so these six category-type features pass the test, and can be used as the features for predicting users' purchasing behavior model.

Table 1: The Card Square of the type characteristics is inspected

Variable	Card Square	Freedom	P
PV	1365.32	1	0.00
Cart	23012.32	1	0.00
Fav	2963.36	1	0.00
Consumer	1063.32	1	0.00
Merchandise	1665.63	1	0.00
Time	1863.24	1	0.00

III. A. 4) Balancing positive and negative samples

Because the original data set purchasing behavior accounts for less than 1%, so the positive and negative samples are extremely unbalanced, the ratio of positive and negative samples in the first 5 million samples is close to 1:50, which will seriously affect the prediction effect, making the prediction results of the negative samples much better than the prediction results of the positive samples. The group draws out all the positive samples totaling 61300 pieces of data, and randomly selects 10 times the amount of positive sample data from all the negative sample datasets to record the samples. And the positive and negative sample data sets were disrupted and used to train and test the model.

III. B. Logistic regression modeling

This subsection uses the LogisticRegression function in Python to establish a logistic regression model on the data set, with different parameter choices, the predictive effect of the same model is not the same. In this paper, the important parameters in the logistic regression model are selected as follows: the regularization parameter penalty is optional with L1 and L2, L1 is more suitable for the model features very much, this paper selects L2 regularization. Classification mode selection parameter Multi_class choose ovr, that is, one_vs_rest binary logistic regression, the sample will be predicted as two categories. Type weight parameter class_weight is used to label the weight of various categories in the classification model, because the data in this paper is unbalanced data, choose balanced, the class library will calculate the weight according to the training samples, balanced ratio between categories, type sample size more weight is low, the sample size is less weight is high. Algorithm convergence maximum number of iterations max_iter, int type, select the default parameter 100. optimization algorithm selection parameter solver select relatively more suitable for the general data size of lbfgs, which is one of the proposed Newton method, the use of the loss function of the second-order derivatives of the loss function, that is, the Hessian matrix to iteratively optimize the loss function. The random seed random_state is set to 0.

Regarding the division of the training and testing sets, since the users in the existing dataset are obtained from all registered users using random sampling method, they are representative. Therefore, in this paper, the users in the original dataset will be divided by three or seven, and 70% of them will be modeled as the training set, and the remaining part will be used as the test set.

III. C. Forecast results and analysis

This study is to predict whether a user will purchase or not in the coming week based on the user's operation log from November 25 to December 3, 2024, with purchase as "1" and non-purchase as "0", which is ostensibly a binary classification problem, but in actual prediction it does not It is ostensibly a binary classification problem, but in the actual prediction, it is not simply predicting the purchase or non-purchase, but giving the probability of the user's purchase, so that the threshold can be set according to the actual situation of the merchant.

Each point on the ROC curve represents the prediction result of the model under a certain threshold, and the whole curve macroscopically shows the overall prediction effect of the model. Using the above mentioned parameters to build a logistic regression model on the training set, the ROC curve predicted on the test set is shown in Figure 3. It can be seen that the curve is skewed to the upper left, and the AUC value of the area under the curve is 0.71362, which is obviously larger than 0.5, indicating that the model's prediction is effective.

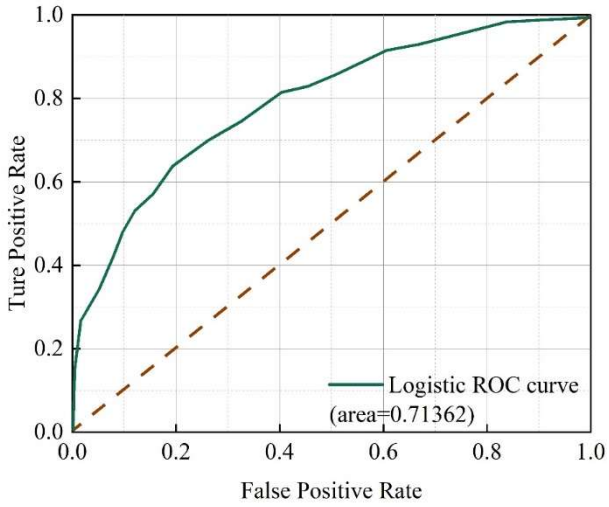

Figure 3: Logistic regression model ROC curve

The ROC curve is a macroscopic display of the model's prediction effect, which is not fine enough for the prediction results under each specific threshold. Therefore, this paper compares the KS curve and the F1 value, which is a comprehensive consideration of recall and accuracy, as two different threshold selection criteria, and determines the threshold value by comprehensively considering the costs incurred by merchants in the process of publicizing users with high purchase intention and the publicity range that needs to be satisfied in order to obtain profit.

F1 value as a standard, when the threshold value of 0.8, the logistic regression model of the F1 value reaches a maximum of 0.35. That is, if the predicted probability of purchase of the user is greater than 0.8, the user will be judged as will be purchased in the coming week, if the probability is less than 0.8, will not be purchased. At this point the logistic regression prediction results are shown in Table 2. It can be seen that at a threshold of 0.8, the recall of the logistic regression model for user purchases is 0.27, i.e., out of all the real purchases, the model predicts 27% of them. The accuracy is 0.48, i.e., of the users predicted to have a greater than 80% probability of purchasing, 48% of them made a real purchase.

Table 2: The threshold value is 0.8 and the logical regression model is predicted

	Precision	Recall	F1-score	Support
0	0.92	0.94	0.98	10932
1	0.48	0.27	0.35	955

Taking the KS value as the standard, the threshold value that makes the KS value the largest is selected, and the KS curve of the logistic regression model is shown in Figure 4. With the increase of the threshold value, the KS curve first rises and then falls, and the model has the strongest classification ability when the threshold value is 0.2236, and the difference between the rate of true cases and the rate of false positive cases is the largest, when the KS value is 0.4362.

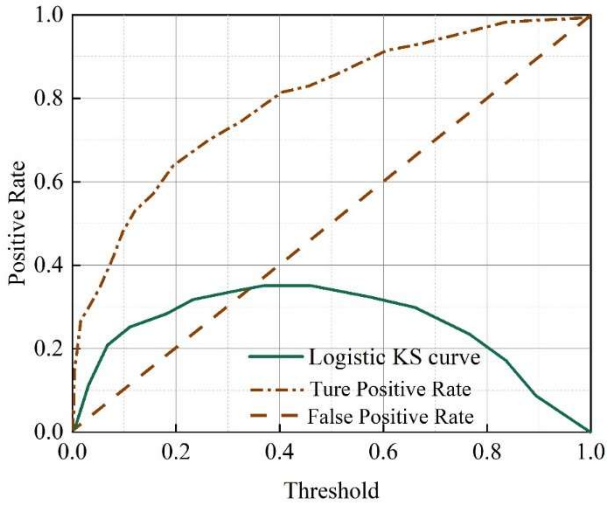

Figure 4: Logistic regression model KS curve

The prediction results of the model at the maximum value of KS are shown in Table 3, and the recall of the logistic regression model for user purchases at the threshold value of 0.2236 is 0.96, i.e., the model predicted 96% of all users who made real purchases, which is significantly higher than the number of users predicted at the threshold value of 0.8. However, the accuracy is significantly lower compared to the threshold value of 0.8, with an accuracy of 0.05. In practical applications, if more attention is paid to the proportion of users who make real purchases among the users predicted to make purchases, the F1 value is a better criterion for the selection of the threshold value.

Table 3: The threshold value is 0.2236 and the logical regression model is predicted

	Precision	Recall	F1-score	Support
0	0.94	0.05	0.07	10932
1	0.06	0.96	0.13	955

Table 4: Each prediction section is counted

Fractional segment	Order number	Unordered number	Total number	The number of ordered people	The number of unordered people
0.0-0.1	10	13	23	0.00%	100.00%
0.1-0.2	13	84	97	3.45%	96.55%
0.2-0.3	60	2488	2548	2.35%	97.65%
0.3-0.4	199	4492	4691	4.24%	95.76%
0.4-0.5	160	1968	2128	7.52%	92.48%
0.5-0.6	129	871	1000	12.90%	87.10%
0.6-0.7	86	399	485	17.73%	82.27%
0.7-0.8	60	270	330	18.18%	81.82%
0.8-0.9	77	189	266	28.95%	71.05%
0.9-1	205	135	340	60.29%	39.71%
Total	979	10900	11879	8.24%	91.76%

The model predicts that the probability of purchase is greater than the threshold judged as the purchase of the user, and vice versa for the user who has not purchased, but only the user will be judged as the purchase and will

not buy is too absolute, and can not accurately express the user's willingness to buy, so the need for the probability that the user will happen to buy the behavior of the statistics, in order to more clearly measure the user's willingness to buy.

The statistics of each prediction score band of the logistic regression model are shown in Table 4. It can be seen that the higher the score band, the greater the proportion of the number of people who buy the score band, to 0.9 ~ 1 score band, for example, the prediction probability of the users in this score band, of which 60.29% of the users have occurred real purchase behavior, indicating that the model prediction effect is good.

The purchase probability value of the logistic regression model prediction result is segmented, and the statistics of the percentage of the number of people placing orders in each fractional segment are shown in Figure 5. It can be seen that the percentage of the number of people purchasing in the fractional segment increases with the increase of the predicted probability value, and the number of people who have not purchased decreases with the increase of the probability value of the fractional segment, and the prediction result is in line with the user's actual purchase.

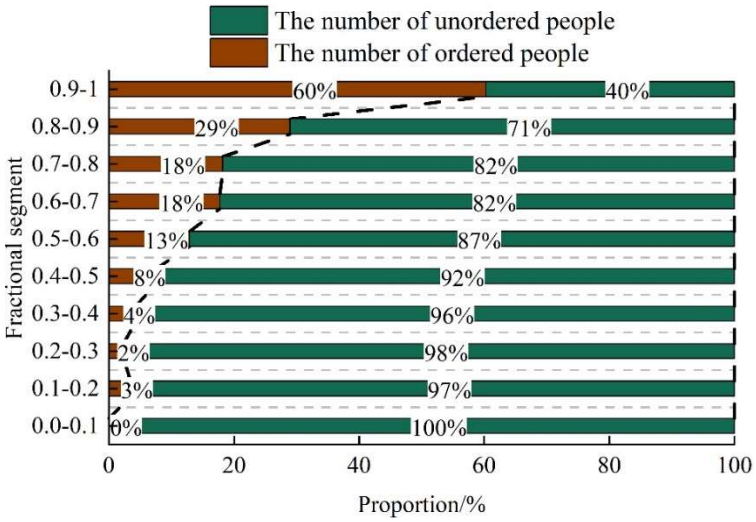

Figure 5: The logical regression model predicts the proportion of the number of points

In summary, after inputting the preprocessed data and category-based features into the constructed logistic regression model, the model can accurately predict the user's purchase line.

IV. Conclusion

In order to improve the prediction performance of user purchasing behavior, the study proposes a method for predicting e-commerce user purchasing behavior based on logistic regression model. Through the chi-square test, click, add to cart, collection, consumer feature group, product feature group and time feature are selected as the features of the model for predicting user purchase behavior.

The experimental results show that the prediction thresholds are selected with the F1 value and KS value as the standard, and comparing the differences of different threshold selection criteria, it is found that the prediction results obtained with the thresholds selected with the F1 value as the standard have a higher checking rate, which is conducive to providing the prediction accuracy. The AUC value of the logistic regression model is 0.71362, which is greater than 0.5, indicating that the logistic regression model constructed in this paper can effectively predict e-commerce users' purchasing behavior.

Funding

1. Discipline Co construction Project of Philosophy and Social Sciences in Guangdong Province's 13th Five Year Plan for 2020 (Project Code: GD20XJY30).

2. Discipline Co construction Project of Philosophy and Social Sciences in Guangdong Province's 14th Five Year Plan for 2022 (Project Code: GD22XGL44).

References

[1] Wongsunopparat, S., & Deng, B. (2021). Factors influencing purchase decision of Chinese consumer under live streaming E-commerce model. Journal of Small Business and Entrepreneurship, 9(2), 1-15.

[2] Wang, Y., Yu, Z., Shen, L., & Dong, W. (2021). E-commerce supply chain models under altruistic preference. Mathematics, 9(6), 632.
[3] Tian, L., Vakharia, A. J., Tan, Y., & Xu, Y. (2018). Marketplace, reseller, or hybrid: Strategic analysis of an emerging e-commerce model. Production and Operations Management, 27(8), 1595-1610.
[4] Wang, J., Dong, K., Sha, Y., & Yan, C. (2022). Envisaging the carbon emissions efficiency of digitalization: The case of the internet economy for China. Technological Forecasting and Social Change, 184, 121965.
[5] Zhang, Z., & Nuangjamnong, C. (2022). The impact factors toward online repurchase intention: A case study of Taobao e-commerce platform in China. International Research E-Journal on Business and Economics, 7(2), 35-56.
[6] Gulfraz, M. B., Sufyan, M., Mustak, M., Salminen, J., & Srivastava, D. K. (2022). Understanding the impact of online customers' shopping experience on online impulsive buying: A study on two leading E-commerce platforms. Journal of Retailing and Consumer Services, 68, 103000.
[7] Qiu, J., Lin, Z., & Li, Y. (2015). Predicting customer purchase behavior in the e-commerce context. Electronic commerce research, 15, 427-452.
[8] Huseynov, F., & Özkan Yıldırım, S. (2019). Online consumer typologies and their shopping behaviors in B2C e-commerce platforms. Sage Open, 9(2), 2158244019854639.
[9] Yao, H. (2022). Analysis Model of Consumer Sentiment Tendency of Commodities in E-Commerce. Frontiers in Psychology, 13, 887923.
[10] Handoyo, S. (2024). Purchasing in the digital age: A meta-analytical perspective on trust, risk, security, and e-WOM in e-commerce. Heliyon, 10(8).
[11] Cacheda, F., Carneiro, V., Fernández, D., & Formoso, V. (2011). Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. ACM Transactions on the Web (TWEB), 5(1), 1-33.
[12] Azad, M. S., Khan, S. S., Hossain, R., Rahman, R., & Momen, S. (2023). Predictive modeling of consumer purchase behavior on social media: Integrating theory of planned behavior and machine learning for actionable insights. Plos one, 18(12), e0296336.
[13] Hála, M., Cvik, E. D., & Pelikánová, R. M. (2022). Logistic regression of Czech luxury fashion purchasing habits during the Covid-19 pandemic–old for loyalty and young for sustainability?. Folia Oeconomica Stetinensia, 22(1), 85-110.
[14] Tang, L., Wang, A., Xu, Z., & Li, J. (2017). Online-purchasing behavior forecasting with a firefly algorithm-based SVM model considering shopping cart use. Eurasia Journal of Mathematics, Science and Technology Education, 13(12), 7967-7983.
[15] Salamzadeh, A., Ebrahimi, P., Soleimani, M., & Fekete-Farkas, M. (2022). Grocery apps and consumer purchase behavior: application of Gaussian mixture model and multi-layer perceptron algorithm. Journal of Risk and Financial Management, 15(10), 424.
[16] Li, J., Pan, S., & Huang, L. (2019). A machine learning based method for customer behavior prediction. Tehnički vjesnik, 26(6), 1670-1676.
[17] Hu, X., Yang, Y., Zhu, S., & Chen, L. (2020, May). Research on a hybrid prediction model for purchase behavior based on logistic regression and support vector machine. In 2020 3rd international conference on artificial intelligence and big data (ICAIBD) (pp. 200-204). IEEE.
[18] Xiao, S., & Tong, W. (2021). Prediction of user consumption behavior data based on the combined model of TF-IDF and logistic regression. In Journal of physics: conference series (Vol. 1757, No. 1, p. 012089). IOP Publishing.
[19] Arun Vashista,Muskaan Arora & Sohela Malik. (2023). E-commerce Environment on Consumers' Traditional Shopping Behavior. Journal of Global Economy, Business and Finance,5(6),
[20] Mahesh K. Singh. (2024). Identification of Speaker from Disguised Voice Using MFCC Feature Extraction, Chi-Square and Classification Technique. Wireless Personal Communications,(prepublish),1-15.
[21] Ali Fallah Tehrani & Diane Ahrens. (2016). Enhanced predictive models for purchasing in the fashion field by using kernel machine regression equipped with ordinal logistic regression. Journal of Retailing and Consumer Services,32,131-138.