Determinants of Sustained Social Interactions on E-commerce Shopping Guide Platforms: A Machine Learning Perspective

Kangdong Ling[1,a], Chenxi Hu[2,*]

[1]Hefei University, Hefei, China
[2]Key laboratory of Financial Big Data, Hefei University, Hefei, China

[a]lingkangdong2020@163.com, [*]hucx@hfuu.edu.cn

Abstract. This study investigates user behavior on e-commerce shopping guide platforms, specifically exploring the determinants of continuous social interactions. Utilizing the XGBoost algorithm, we developed a predictive model incorporating sentiment tendencies derived from user comments. By dividing a dataset of 3,212 ZHI-TECH users into training, validation, and testing subsets, the model was optimized through AutoGluon-Tabular. The optimized model was then applied to the testing dataset to predict user engagement in social interactions. With an AUC of 0.7311, Precision of 0.7028, Recall of 0.9757, and F1 score of 0.8171, the model exhibited a strong predictive performance.

Our findings provide e-commerce platforms with valuable insights into fostering user engagement, offering potential strategies to enhance user activity and satisfaction. These strategies include promoting user interaction, enhancing personal brand value, enriching community content, monitoring users' emotional tendencies, and addressing the needs of VIP users. The study contributes significantly to the academic understanding of social interaction behaviors on e-commerce platforms.

Keywords: E-commerce Guide Platforms; Machine Learning; social interactions; XGBoost; GPT-3

1 Introduction

Social e-commerce is a business model that merges the interactive features of social networks with the transaction functionality of e-commerce. In this model, users' social behaviors are closely connected to their shopping experiences, and they interact with others through social activities such as liking, commenting, and sharing. Meanwhile, users' purchasing decisions can also be influenced by social activities, which in turn can change their shopping decisions[1]. This interaction not only enhances the user shopping experience but also provides merchants with rich user behavior data and market insights[2]. With the rise of social e-commerce, continuous social interaction among users has become a key factor influencing brand loyalty and purchase

Y. Kuang et al. (eds.), *Proceedings of the 2024 5th International Conference on Education, Knowledge and Information Management (ICEKIM 2024)*, Atlantis Highlights in Computer Sciences 22,
https://doi.org/10.2991/978-94-6463-502-7_55

intention. Therefore, understanding and analyzing user interactive behaviors on social e-commerce platforms is of great significance for improving the user shopping experience and enhancing brand business value[3].

The primary objective of this paper is to employ the XGBoost model and Natural Language Processing (NLP) to forecast users' ongoing social interaction behavior on e-commerce shopping guide platforms.

2 Literature Review

2.1 E-Commerce Shopping Guide Platform

E-commerce shopping guide platform is a platform that provides consumers with shopping information. These platforms usually collect and organize product information from e-commerce platforms and provide functions such as price comparison and discount information to help consumers make more informed shopping decisions.

E-commerce shopping guide platform help consumers save money in three ways: coupons, cashback, and earning opportunities through shared promotions. They act as middlemen, finding you deals and comparing prices across platforms like Taobao, Tmall, JD.com, and Pinduoduo. This expands your online shopping options, letting you save and even earn money[4].

For e-commerce shopping guide platforms, boosting user loyalty necessitates reducing product selection costs and enhancing the quality of user decision-making. To achieve this, these platforms must increase the accuracy and novelty of product recommendations while curtailing misinformation and information overload[5]. The quality of information, services, and systems positively influences consumers' continued platform usage intention. Furthermore, user satisfaction with e-commerce shopping guide platforms is a crucial predictor of users' ongoing platform usage. The services, social features, and social interactions facilitated by these platforms influence user satisfaction[6].

2.2 Beijing Zhidemai Technology (ZHI-TECH)

Beijing Zhidemai Technology Co., Ltd. (ZHI-TECH) is a leading e-commerce shopping guide platform in China, founded in 2011 and listed in 2019. The platform focuses on user-generated content, with over 70% of its customers acquired through this method in 2021. ZHI-TECH users are known for their active social interactions, making it an ideal platform to study factors influencing sustained user engagement[7].

2.3 Generative Pre-trained Transformer 3 (GPT-3)

On e-commerce shopping guide platforms, the sentiment orientation of user reviews significantly impacts the decision-making process and interaction levels of other users[8]. GPT-3 serves as a tool for sentiment analysis of review texts, classifying reviews into positive, neutral, or negative categories. By scrutinizing an extensive amount of review data, we can establish an average sentiment orientation of an individual user when they compose reviews within these three categories. Utilizing these sentiment averages as input variables enables the construction of a model of sustained user social interaction. Integrating the sentiment analysis capabilities of GPT-3, we can predict user social interaction behavior on the platform with higher accuracy, thereby providing more effective personalized services and recommendations for the platform[9]. Ultimately, this facilitates the promotion of sustained user interaction on e-commerce shopping guide platforms and enhances the overall user experience.

2.4 eXtreme Gradient Boosting (XGBoost)

XGBoost is an efficient machine learning model grounded in gradient boosting algorithms, proposed by Chen. By leveraging parallel computing and feature selection techniques to expedite the training process, XGBoost has demonstrated remarkable predictive capabilities in various competitions and practical applications. Furthermore, XGBoost exhibits superb scalability, enabling it to process large-scale datasets, thereby making it an ideal choice for tackling complex problems.

In the context of e-commerce shopping guide platforms, XGBoost has proven to be an effective model. For instance, used the XGBoost model to predict user online shopping behavior[10], while Lien et al. (2020) combined XGBoost and skip-gram models to predict the popularity of online reviews. In this study, we draw upon these previous research findings and apply XGBoost to predict user sustained social interaction one-commerce shopping guide platforms.

3 Research Design

The data in this article uses NoxPlayer emulator and Mitmproxy to obtain user data on the ZHI-TECH app. After the data crawling phase, and subsequent removal of errors and duplicate entries, we have amassed a total of 8,428 user personal information records, 3,038,668 user social interaction records, and 792,320 user comment records. We perform data preprocessing based on the collected data, implement discriminant models based on four machine learning algorithms, evaluate their performance, and apply the test samples to the algorithm with the best discriminant validation.

3.1 Data Preprocessing

In order to calculate the sentiment value of review text, We fine-tuned GPT-3 to analyze sentiment in e-commerce reviews. By training on 2,000 labeled examples, we built a model that assigns sentiment scores (based on log probabilities) to user reviews. These scores are then averaged to create a user's overall sentiment tendency.

By analyzing the characteristics of user information and referring to the theoretical definitions of user activities on social e-commerce platforms by Elnaz Meydani et al.[11], we divided the independent variables into three layers: social identity (individual level), social interaction (Conversation level), and social interaction (community level), and identified a total of 14 predictive factors. To determine whether a user has continuous social interaction, we checked if the interval between the data acquisition date and the user's last social interaction date was greater than one month. The independent variable data and dependent variable are shown in Table 1.

Table 1. Description of variant

Layer	Category
Social Identification (Individual layer)	Number of medals
Social Interaction (Conversation Layer)	Number of comments
Social Interaction (Community layer)	Number of fans
Social Interaction (Community layer)	Number of following
Social Interaction (Community layer)	Number of visits
Social Identification (Individual layer)	Whether to edit the personal profile
Social Identification (Individual layer)	Length of personal profile
Social Interaction (Conversation Layer)	Number of Explosives
Social Identification (Individual layer)	VIP level
Social Identification (Individual layer)	Length of name
Social Interaction (Conversation Layer)	Positive Sentiment Value
Social Interaction (Conversation Layer)	Neutral Sentiment Value
Social Interaction (Conversation Layer)	Negative Sentiment Value
Social Interaction (Conversation Layer)	Number of Experiences
	Whether to continue social interaction

Social identity activities, classified at the individual level, primarily involve the creation of personalized information such as personal details, tags, and badges. Social interaction activities, at the conversational level, revolve around the exchange of information and opinions between users, including actions like liking, commenting, and voting. Lastly, social interaction at the community level focuses on establishing and maintaining social relationships, with actions such as following other users and participating in discussion groups. Each layer provides a unique perspective and contributes to the overall user engagement in a social e-commerce platform.

3.2 Creating Discriminant Models

We developed classification models using XGBoost, Random Forest, SVM, Logistic Regression, and k-NN with AutoGluon, an AutoML framework for model ensembling and stacking.To ensure model reliability, we employed 10-fold cross-validation alongside AutoGluon's optimization techniques for various models. Cross-validation evaluates model generalizability, which is crucial for accurate predictions.

In essence, we trained multiple models, validated their performance on unseen data, and selected the best performing model using AutoGluon's hyperparameter tuning and ensembling capabilities.

4 Evaluation and Selection of Discriminant Model

4.1 Discriminant Model Performance Validation Indicator

This study prioritizes classification accuracy to evaluate models. We choose the model with the highest accuracy and use variable selection to assess new variables. We then use precision and recall, calculated from a confusion matrix, to analyze each model's performance further. Finally, the F1-score, combining precision and recall, provides a single metric for model comparison.

4.2 Receiver Operating Curve (ROC) of a Discriminant Model

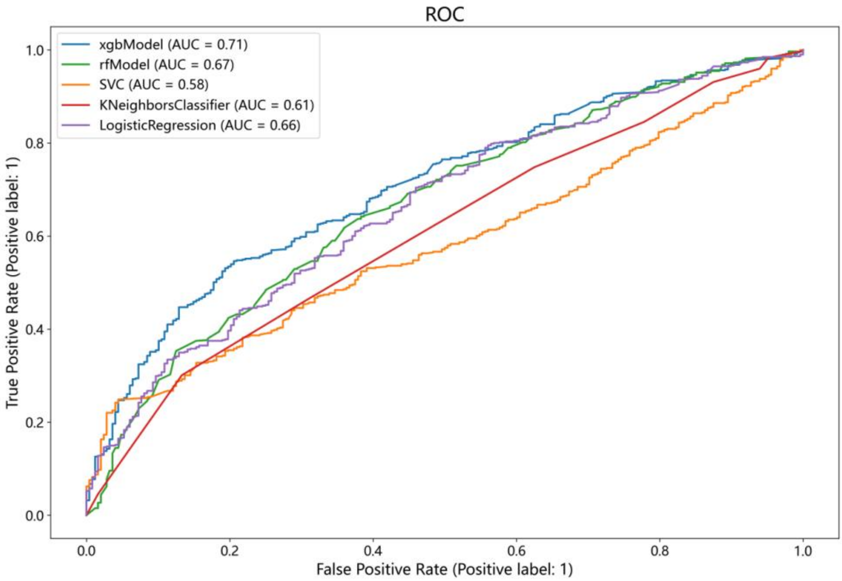

Fig. 1. ROC

A model's accuracy is indicated by higher TP values and lower FP values, which result in an upward convex ROC curve. The area under the ROC curve, known as the AUC, ranges between 0.5 and 1. The ROC analysis results are depicted in Figure 1. Among the models, XGBoost yielded the highest AUC value of 0.71, followed by Random Forest with 0.67, Logistic Regression with 0.66, SVM with 0.58, and KNN with 0.61. However, considering the superior accuracy, reproducibility, and F1 score of the XGBoost model in this study, it exhibits superior classification validation capabilities compared to the other models, as illustrated in Table 2. Therefore, we elected to use the XGBoost algorithm for building the classification model.

Table 2. Evaluate the predictive power of the Model

Model	AUC	F1-Score	Precision	Recall
XGBoost	0.7064	0.5877	0.6294	0.5848
RF	0.6682	0.5705	0.6163	0.5712
SVM	0.5805	0.4138	0.3529	0.5000
LR	0.6648	0.4138	0.3529	0.5000
KNN	0.6082	0.5289	0.5484	0.5336

5 Discriminant Model Testing and Analysis

5.1 Model Optimization Using AutoGluon

Upon establishing our dataset divisions and the implementation of the XGBoost classification model, our next step was model optimization. We opted for AutoGluon-Tabular, an open-source AutoML framework known for its efficiency and robustness in training highly accurate models on raw tabular data. Its distinctive approach, ensembling multiple models and stacking them in layers, has proven to be more effective than the pursuit of the single best model. Additionally, comprehensive evaluations have shown AutoGluon to be faster, more robust, and notably more accurate than other AutoML platforms. As such, using AutoGluon to optimize our XGBoost model aligned with our objective of augmenting the model's performance and robustness while keeping abreast with the current best practices in data science.

5.2 Prediction Using XGBoost Discriminant Model

For further discussion, this study employs the testing dataset to predict whether ZHI-TECH users will continue social interactions based on the optimized XGBoost classification model. Figure 2 presents the binary classification results for the testing dataset. Among the 843 ZHI-TECH users in the testing samples,577 individuals who did continuously engage in social interactions, of which 14 were erroneously identified as users who did not continuously engage in social interactions. Out of the 266 users who did not continuously engage in social interactions, 28 were correctly identified, and 238 were incorrectly identified. The results displayed in Table 3 are as follows: AUC = 0.7311, Precision = 0.7028, Recall = 0.9757, F1 = 0.8171. In summary, our optimized XGBoost classification model demonstrates commendable performance in predicting the continuity of social interactions among ZHI-TECH users.

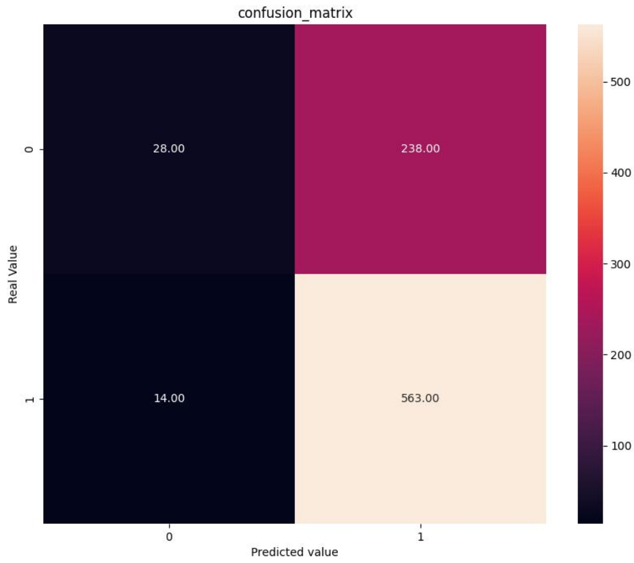

Fig. 2. Confusion Matrix

Table 3. Evaluation of XGBoost model

Model	AUC	F1	Precision	Recall
XGBoost	0.7311	0.8171	0.7028	0.9757

5.3 Evaluation of Variables Affecting Sustained Social Interaction

After making predictions using validation data, we further analyzed the collected variables (a total of 13 variables) to determine more information and identify ongoing social interactions. To explain the effect of variables on the prediction results, we introduced the SHAP (SHapley Additive exPlanations) values. SHAP is an additive interpretation model developed by Lundberg in 2017, inspired by cooperative game theory. At its core, the Shap Values for each feature are calculated to reflect the extent to which the feature contributes to the predictive power of the Model as a whole. Shap interprets the predictive value of the Model as the sum of the Shap Values of each input feature.The SHAP value can reflect the impact of each feature on the final prediction value and can show the positivity or negativity of the impact, increasing the interpretability of the Model. we selected a sample of 500 user data instances that were not included in the model training process. The SHAP values were calculated for these instances and visualized in a Summary Plot, as depicted in Figure 3.

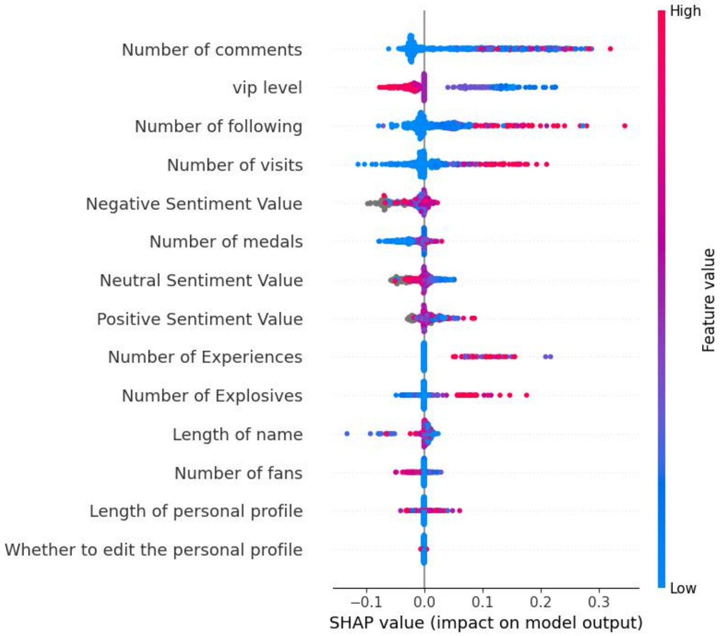

Fig. 3. Summary Plot

The vertical arrangement of the variables in the figure illustrates their relative contribution to the prediction. The horizontal axis of the data points signifies the impact of the feature value on the output, while the color indicates the magnitude of the feature value. As can be gleaned from the figure, factors such as the number of visits, comments, VIP level, and followers significantly influence the output value.

From Figure 3, it can be observed that users who frequently engage in activities such as commenting, visiting, and gathering followers are more inclined to sustain social interactions. Interestingly, a counterintuitive trend is noticed among users with a higher VIP level, who appear more likely to discontinue social interactions. Another notable trend is related to sentiment values; users with a high neutral sentiment value tend to be less likely to maintain social interactions. This factor seems to be inversely related to positive and negative sentiment values. Although the overall effect is relatively subdued, users with higher positive and negative sentiment values demonstrate a higher propensity for continuous social interactions.

5.4 Analysis of the Impact of Users' Continuous Social Interaction

This study investigates factors influencing user engagement on a social platform. The analysis finds that users who actively comment, visit others' pages, and gather followers are more likely to maintain social interactions. Interestingly, users with a high VIP level and those expressing mostly neutral sentiment are less likely to stay engaged.

Active users: Frequent commenting, visiting, and follower gathering indicate user activity and willingness to interact. This fosters connections and motivates continued participation.

Popularity: A high number of visits suggests user appeal, potentially due to content or personal brand. This motivates users to maintain engagement to retain their influence.

Following: A larger following indicates user interest in the community and a desire to interact for information, connection, or support. This focus on the community fuels continued engagement.

Sentiment: While the effect is subtle, users with a more positive or negative sentiment are slightly more likely to stay engaged than those expressing mostly neutral sentiment. Positive interactions likely provide a feedback loop that motivates continued participation. Negative interactions might lead to short-term engagement, but negativity can ultimately hinder the social experience.

Recognition: The number of medals a user has correlates with a higher likelihood of continued social interaction. Earning medals suggests user engagement and a desire for community recognition. This recognition might also attract attention from others, further stimulating interaction.

VIP level: Surprisingly, users with a higher VIP level are less likely to stay engaged. This could be because these users have already established their presence and may seek new experiences elsewhere. Conversely, new users with lower VIP levels might be more actively engaged in building their social network, explaining their higher persistence.

In conclusion, this study reveals that active participation, popularity within the community, and emotional expression (both positive and negative) are all positive indicators of a user's likelihood to maintain social interaction on the platform.

6 Conclusions

This study has revealed some insightful findings about user behavior on e-commerce shopping guide platforms. By utilizing the XGBoost algorithm and incorporating sentiment tendencies from comments into the model, we have been able to successfully predict whether users will continue to engage in social interactions on these platforms. The results have highlighted some critical factors driving user engagement.

As the study reveals, the number of comments and homepage visits are paramount factors influencing the continuous engagement of users. This suggests that user interactions play a significant role in fostering sustained activity on the platform. For e-commerce shopping guide platforms, it may therefore be beneficial to implement measures that promote a vibrant, interactive atmosphere within the community, as well as strengthening user interactions.

The study also brings to light the importance of the community layer of social interactions, which includes the number of followers a user has and the number of users they follow. Interestingly, these two aspects have different impacts on a user's ongoing engagement. Following more users indicates an interest in their content and a desire to receive more in the future, often leading to more active engagement.

The importance of social identification is also highlighted in the study, with variables like the number of medals showing a positive impact on continuous social interactions. These accolades, representing a user's past engagement and achievements within the community, serve as motivators for ongoing participation.

Overall, the findings from this study offer valuable insights for e-commerce shopping guide platforms, providing potential strategies to enhance user engagement and activity. However, it is important to note that while these findings are relevant to the context of the study, different platforms might have different dynamics and factors driving user engagement, and the models might need to be adjusted accordingly.

References

1. Xq L, Dashuai R, Zhen J. (2021)Influence of social interaction on the purchase behavior of social commerce consumers. Human Systems Management, 40: 917-35. 10.3233/HSM-190801
2. Huang N, Sun T, Chen P-y, Golden J. (2017)Social media integration and e-commerce platform performance: A randomized field experiment. https://dx.doi.org/10.2139/ssrn.2969670
3. Xiang L, Zheng X, Lee MKO, Zhao D. (2016)Exploring consumers' impulse buying behavior on social commerce platform: The role of parasocial interaction. International Journal of Information Management, 36: 333-47. https://doi.org/10.1016/j.ijinfomgt.2015.11.002
4. Du M. (2019)Preferential Purchase of E-commerce Model in China. Economics and Management Science, 1: 1-6. 10.36012/ems.v1i1.956
5. Qi Y, Changqi C. (2017)A Study on the Influence of Online Tips for E-commerce Recommended Products on Consumer Willingness to Buy. Contemporary economic management, 39: 22-30. 10.13253/j.cnki.ddjjgl.2017.11.004
6. Zhang S, Liu L, Li M, Tao Q, Zhang R, Xia Y. Understanding Continuance Usage Intention of Shopping Guide Apps in Social Commerce. International Conference on Human-Computer Interaction: Springer; 2021. p. 414-21. https://doi.org/10.1007/978-3-030-90179-0_53
7. Yanxin L. A Study on the Profit Model of the ZHI-TECH and its Financial Evaluation Beijing Institute of Graphic Communication; 2021. https://doi.org/10.1007/978-3-030-92836-0
8. Pang B, Lee L. (2008)Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 2: 1-135. 10.1561/1500000011
9. Paul J, Ueno A, Dennis C. ChatGPT and consumers: Benefits, Pitfalls and Future Research Agenda. International Journal of Consumer Studies, n/a: https://doi.org/10.1111/ijcs.12928
10. Liu C-J, Huang T-S, Ho P-T, Huang J-C, Hsieh C-T. (2020)Machine learning-based e-commerce platform repurchase customer prediction model. PLOS ONE, 15: e0243105. 10.1371/journal.pone.0243105
11. Meydani E, Düsing C, Fedrau A, Trier M. (2022)The "black box" of social commerce platforms-a closer look at users' activities. 10.3233/HSM-190801