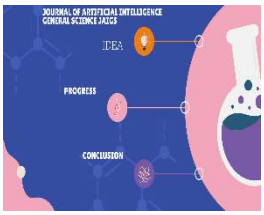
Vol., 5 Issue 01, June, 2024
Journal of Artificial Intelligence General Science JAIGS

https://ojs.boulibrary.com/index.php/JAIGS



Prediction of financial customer buying behavior based on machine learning

Neha Romanenko1*,Kritika Sharma2, Siddharth Verma3

1 Financial Risk, Indian School of Business (ISB), Hyderabad
2Business administration,Indian Institute of Management (IIM) Bangalore

3 Electronic information engineering,Indian Institute of Technology (IIT) Kanpur

*Corresponding author E-mail:Gupta889@gmail.com

ARTICLEINFO
Article History:
Received:
01.05.2024
Accepted:
25.05.2024
Online: 25.06.2024
Keyword: Feature Engineering, Machine Learning Algorithms, Predictive Modeling, Customer Behavior Analysis

ABSTRACT

This study investigates the prediction of online purchasing behavior on the palm life APP through comprehensive analysis of customer operation logs, attribute sets, and purchase labels. Utilizing advanced feature engineering techniques, including time-based metrics, frequency analyses, and category-specific operations, we constructed a robust feature system comprising 61 associated features. The predictive models, leveraging Logistic Regression and LightGBM algorithms, were evaluated using cross-validation and AUC scores, demonstrating strong generalization capabilities and effectiveness in predicting customer behavior. Findings highlight the significance of personalized recommendations and targeted marketing strategies in enhancing customer engagement and optimizing operational efficiencies for e-commerce platforms. This research contributes to both theoretical advancements in consumer behavior prediction and practical implications for enhancing customer experience and service personalization in fintech applications.

1. INTRODUCTION

In the context of the rapid development of financial technology, e-commerce platforms are faced with the challenge of how to effectively deliver product information and accurately serve customers. This study aims to predict the online purchasing behavior of customers by analyzing the behavior data of customers buying coupons in the palm life APP. Specifically, this paper presents an innovative method and research idea, which lays a foundation for the future research of similar prediction problems.

Firstly, we introduce the customer operation log data set, customer attribute set and purchase label set, analyze the data structure in detail, and adopt appropriate methods to deal with outliers. Secondly, through descriptive statistical analysis, the shopping habits of customers are revealed, and the characteristics are mined from multiple perspectives such as time, frequency, interval, module and operation mode. In the feature engineering phase, the research team expanded to 61 associated features and combined them with customer attributes to build a complete feature system.

In order to build an effective prediction model, this study uses advanced machine learning algorithms such as XGBoost and LightGBM for training and learning. Cross-validation and AUC scores are used to evaluate the generalization ability of the model to ensure that the model is good at predicting unknown data.

Finally, combined with the results of characteristic analysis, this paper summarizes the main results of the research, and puts forward suggestions to improve customers' coupon purchasing behavior. These suggestions can not only help e-commerce platforms achieve personalized and customized services, improve the purchase rate of multiple categories of goods, but also help merchants optimize inventory management, reduce costs, promote the development of stores, and improve the shopping experience of customers.

To sum up, this study not only provides a new perspective and method for the prediction of e-commerce consumer behavior in theory, but also demonstrates its application and effect in practice, providing a valuable reference for further research and practice in related fields.

2. RELATED WORK

2.1 Finance related concepts

1. Bank APP customer relationship management theory

The research shows that consumers' shopping intention shows different rules at different time points. There are some differences in consumers' shopping intention between the first purchase and multiple purchases, which provides a basis for the establishment of a customer purchasing behaviour model. Based on the purchase log data of the customers of the APP, the paper puts forward the serial characteristics and time characteristics to describe the customers' purchase behaviour. It emphasises

the importance of personalized recommendations and detailed coupon information to improve purchase efficiency.

2. Consumption theory of online purchase

- Online purchase decisions include demand arousal, information collection, selection evaluation, purchase decision and post-purchase evaluation. The advantages of online consumption include rich variety of goods, price transparency and convenience, and price psychology and preference are important factors. The promotional effect of coupons on consumers is related to price preference. The theory of consumer behavior provides guidance for the formulation of preferential strategies.

2.2 Computer-related concepts

1. LightGBM and XGBoost algorithm

- LightGBM: Decision tree-based gradient lifting algorithm with efficient training speed and optimization strategies, such as feature parallelism and histogram acceleration, further improves model effectiveness and training efficiency by introducing new technologies such as GOSS and EFB.

- XGBoost: It is also an optimized distributed library based on gradient lift tree, which realizes efficient machine learning algorithm through the promotion of parallel decision trees, and gradually optimizes the prediction results of each decision tree to obtain accurate prediction models.

2. AUC value

- AUC (Area under the curve) is used to evaluate the performance of the binary classification model. The ROC curve shows the true case rate (TPR) and false positive case rate (FPR) of the model under different thresholds. The closer the AUC value is to 1, the better the model performance.

These concepts and technologies are interwoven in research and application, providing an important theoretical basis and technical support for understanding and predicting consumer behavior and optimizing financial services.

3. Methodology

In this article, the research team explored the predictive analysis of customers' online purchasing behavior from the data of the customer purchase prediction contest under the consumer finance scenario held by the Credit Card Center of China Merchants Bank. The following are the main methodological parts:

3.1 Data introduction and processing

1. Data set description:

Data set 1: Including personal attributes and consumption data of 80,000 credit card customers, a total of 30 columns, mainly used to analyze the relationship between customer attributes and credit card consumption.

Data set 2: APP customer operation behaviour logs, which record the operation behaviours in March 2021, are used to analyse the behaviour patterns and preferences of customers on the APP.

Data set 3: Label data that contains information on whether customers will purchase coupons in the coming week.

3.2 Data processing:

Desensitize and standardise personal attributes and credit card consumption data to prepare for subsequent analysis of customer operation behavior log data.

Split and aggregate the fields of APP customer operation behavior logs, extract time features such as date, hours and weeks, and statistically analyze the number and click rate of customers on different modules.

The clustering algorithm is used to analyse customer behavior patterns and preferences, such as customer activity, time interval, number of actions and category operations.

3.3 Feature Engineering

Firstly, time interval feature extraction is one of the important components of feature engineering. By analyzing the customer's operation record time stamp on the APP, we calculate the average, maximum, minimum, standard difference and other statistics of each customer's operation record time interval. These characteristics not only reflect the customer's operating frequency and stability, but also reveal the distribution law of the customer's active time. For example, a short average interval may indicate frequent usage habits, while a long maximum interval may indicate irregular patterns of behavior, information that is critical to understanding and predicting customer purchase intentions.

Secondly, the construction of behavior frequency features can further deepen the understanding of customer behavior patterns. By counting the average number of behaviours and the date of behaviours, we reveal the frequency and time tendency of customers to engage in behaviours in a specific period of time. These characteristics can not only help identify active users, but also help to discover the underlying patterns of different behavior patterns, providing the basis for personalized recommendations and marketing strategies.

The introduction of continuous active features further enriches the feature system. By analyzing continuous customer activity, such as the frequency and pattern of APP launches multiple times a day, we can gain insight into customer engagement and the stability of behavior patterns. These characteristics are critical for predicting long-term customer engagement and loyalty, helping to identify behavioral patterns that are important for customer conversion and satisfaction.

Finally, the extraction of category operation characteristics can help identify the customer's behavioral preferences and engagement in different categories. We counted the operation times of customers on different modules, such as cate_1, cate_2, cate_3, etc., and revealed the preference and use frequency of customers on different functional modules. These characteristics not only help to understand customers' points of interest and behavioral focus, but also provide data support for personalized recommendations and precision marketing. feature engineering builds a rich and diverse feature system

through in-depth mining and statistical analysis of customer behavior data, which provides a key data basis for subsequent model training and prediction tasks. These features not only help to improve the predictive performance and generalization ability of the model, but also provide substantial support and guidance for data-driven decision making in the fintech field.

3.4 Model Prediction

1. Build the model:

Logistic regression model and Light GBM algorithm based model are selected to predict customer's online purchasing behavior.

Feature correlation analysis and importance screening are used to build a complete feature system and improve the prediction accuracy and generalization ability of the model.

2. Model training and evaluation:

The logical regression model and Light GBM algorithm model were trained with appropriate hyper parameters to evaluate the prediction effect of the model.

The classification performance of the model is evaluated by AUC values to ensure good performance of the model on the data set.

3.5 Conclusion

Through the above methods, the research team can deeply understand the behavior patterns and preferences of customers on the handheld life APP, and provide targeted personalized recommendations and customer classification strategies for CMB. This methodology is not only limited to theoretical analysis, but also combines practical data processing and model application, providing a practical reference for data analysis and prediction in the field of fin tech.

4. Conclusion

Based on the detailed analysis and methodology presented in this study, we have achieved significant insights into predicting online purchasing behavior on the palm life APP. Through comprehensive data analysis, feature engineering, and advanced machine learning techniques such as Logistic Regression and LightGBM, we have successfully built robust predictive models. These models demonstrate strong performance in evaluating and predicting customer behavior, as evidenced by high AUC scores during cross-validation.

Furthermore, our findings underscore the importance of personalized recommendations and precise marketing strategies in enhancing customer engagement and purchase efficiency. By leveraging the rich feature system developed from customer behavior data, including time intervals, behavior frequencies, continuous activities, and category preferences, we can optimize promotional efforts and improve customer satisfaction. These insights not only contribute to academic research in consumer behavior prediction but also provide practical implications for e-commerce platforms aiming to improve service customization and operational efficiency.

In conclusion, this research not only advances theoretical understanding but also demonstrates practical applications in fintech. By refining our understanding of customer preferences and behaviors, we can

pave the way for future innovations in consumer analytics and decision-making processes within the financial technology sector.

5. REFERENCES

[1] Shi, Y., Li, L., Li, H., Li, A., & Lin, Y. (2024). Aspect-Level Sentiment Analysis of Customer Reviews Based on Neural Multi-task Learning. Journal of Theory and Practice of Engineering Science, 4(04), 1-8.

[2] Jiang, W.; Qian, K.; Fan, C.; Ding, W.; Li, Z. Applications of generative AI-based financial robot advisors as investment consultants. Appl. Comput. Eng. 2024, 67, 28–33, https://doi.org/10.54254/2755-2721/67/2024ma0057.

[3] Ding, W., Zhou, H., Tan, H., Li, Z., & Fan, C. (2024). Automated Compatibility Testing Method for Distributed Software Systems in Cloud Computing.

[4] Fan, C.; Li, Z.; Ding, W.; Zhou, H.; Qian, K. Integrating artificial intelligence with SLAM technology for robotic navigation and localization in unknown environments. Appl. Comput. Eng. 2024, 67, 22–27, https://doi.org/10.54254/2755-2721/67/2024ma0056.

[5] Zhan, T.; Shi, C.; Shi, Y.; Li, H.; Lin, Y. Optimization techniques for sentiment analysis based on LLM (GPT-3). Appl. Comput. Eng. 2024, 67, 41−47, https://doi.org/10.54254/2755-2721/67/2024ma0060.

[6] Li, Huixiang, et al. "AI Face Recognition and Processing Technology Based on GPU Computing." Journal of Theory and Practice of Engineering Science 4.05 (2024): 9-16.

[7] Yuan, J., Lin, Y., Shi, Y., Yang, T., & Li, A. (2024). Applications of Artificial Intelligence Generative Adversarial Techniques in the Financial Sector. Academic Journal of Sociology and Management, 2(3), 59-66.

[8] Guo, L., Li, Z., Qian, K., Ding, W., & Chen, Z. (2024). Bank Credit Risk Early Warning Model Based on Machine Learning Decision Trees. Journal of Economic Theory and Business Management, 1(3), 24-30.

[9] Li, Zihan, et al. "Robot Navigation and Map Construction Based on SLAM Technology." (2024).

[10] Fan, C., Ding, W., Qian, K., Tan, H., & Li, Z. (2024). Cueing Flight Object Trajectory and Safety Prediction Based on SLAM Technology. Journal of Theory and Practice of Engineering Science, 4(05), 1-8.

[11] Ding, W.; Tan, H.; Zhou, H.; Li, Z.; Fan, C. Immediate traffic flow monitoring and management based on multimodal data in cloud computing. Appl. Comput. Eng. 2024, 71, 1–6, https://doi.org/10.54254/2755-2721/71/2024ma0052.

[12] Liu, S., Yan, K., Qin, F., Wang, C., Ge, R., Zhang, K., Huang, J., Peng, Y. and Cao, J., 2024. Infrared Image Super-Resolution via Lightweight Information Split Network. arXiv preprint arXiv:2405.10561.

[13] Qian, K., Fan, C., Li, Z., Zhou, H., & Ding, W. (2024). Implementation of Artificial Intelligence in Investment Decision-making in the Chinese A-share Market. Journal of Economic Theory and Business Management, 1(2), 36-42.

[14] Shi, Y.; Yuan, J.; Yang, P.; Wang, Y.; Chen, Z. Implementing intelligent predictive models for patient disease risk in cloud data warehousing. Appl. Comput. Eng. 2024, 67, 34–40, https://doi.org/10.54254/2755-2721/67/2024ma0059.

[15] Lin, Y., Li, A., Li, H., Shi, Y., & Zhan, X. (2024). GPU-Optimized Image Processing and Generation Based on Deep Learning and Computer Vision. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 5(1), 39-49.
[16] Tian, J.; Li, H.; Qi, Y.; Wang, X.; Feng, Y. Intelligent medical detection and diagnosis assisted by deep learning. Appl. Comput. Eng. 2024, 64, 116−121, https://doi.org/10.54254/2755-2721/64/20241356.
[17] Zhan, X., Shi, C., Li, L., Xu, K., & Zheng, H. (2024). Aspect category sentiment analysis based on multiple attention mechanisms and pre-trained models. Applied and Computational Engineering, 71, 21-26.
[18] Wu, B., Xu, J., Zhang, Y., Liu, B., Gong, Y., & Huang, J. (2024). Integration of computer networks and artificial neural networks for an AI-based network operator. Applied and Computational Engineering, 64, 115-120.
[19] Xiao, J., Wang, J., Bao, W., Deng, T. and Bi, S., Application progress of natural language processing technology in financial research.
[20] Liang, P.; Song, B.; Zhan, X.; Chen, Z.; Yuan, J. Automating the training and deployment of models in MLOps by integrating systems with machine learning. Appl. Comput. Eng. 2024, 67, 1−7, https://doi.org/10.54254/2755-2721/67/20240690.
[21] Li, A., Yang, T., Zhan, X., Shi, Y., & Li, H. (2024). Utilizing Data Science and AI for Customer Churn Prediction in Marketing. Journal of Theory and Practice of Engineering Science, 4(05), 72-79.
[22] Wu, B., Gong, Y., Zheng, H., Zhang, Y., Huang, J., & Xu, J. (2024). Enterprise cloud resource optimization and management based on cloud operations. Applied and Computational Engineering, 67, 8-14.
[23] Xu, J.; Wu, B.; Huang, J.; Gong, Y.; Zhang, Y.; Liu, B. Practical applications of advanced cloud services and generative AI systems in medical image analysis. Appl. Comput. Eng. 2024, 64, 83–88, https://doi.org/10.54254/2755-2721/64/20241361.
[24] Zhang, Y.; Liu, B.; Gong, Y.; Huang, J.; Xu, J.; Wan, W. Application of machine learning optimization in cloud computing resource scheduling and management. Appl. Comput. Eng. 2024, 64, 17–22, https://doi.org/10.54254/2755-2721/64/20241359.
[25] Huang, J.; Zhang, Y.; Xu, J.; Wu, B.; Liu, B.; Gong, Y. Implementation of seamless assistance with Google Assistant leveraging cloud computing. Appl. Comput. Eng. 2024, 64, 170−176, https://doi.org/10.54254/2755-2721/64/20241383.
[26] Gong, Y.; Huang, J.; Liu, B.; Xu, J.; Wu, B.; Zhang, Y. Dynamic resource allocation for virtual machine migration optimization using machine learning. Appl. Comput. Eng. 2024, 57, 1−8, https://doi.org/10.54254/2755-2721/57/20241348.
[27] Bi, Shuochen, Wenqing Bao, Jue Xiao, Jiangshan Wang, and Tingting Deng. "Application and practice of AI technology in quantitative investment." arXiv preprint arXiv:2404.18184(2024).
[28] Xu, Jiahao, et al. "AI-BASED RISK PREDICTION AND MONITORING IN FINANCIAL FUTURES AND SECURITIES MARKETS." The 13th International scientific and practical conference “Information and innovative technologies in the development of society”(April 02–05, 2024) Athens, Greece. International Science Group. 2024. 321 p.. 2024.
[29] Wang, Yong, et al. "Machine Learning-Based Facial Recognition for Financial Fraud Prevention." Journal of Computer Technology and Applied Mathematics 1.1 (2024): 77-84.
[30] Song, Jintong, et al. "LSTM-Based Deep Learning Model for Financial Market Stock Price Prediction." Journal of Economic Theory and Business Management 1.2 (2024): 43-50.
[31] Cheng, Qishuo, et al. "Monetary Policy and Wealth Growth: AI-Enhanced Analysis of Dual Equilibrium in Product and Money Markets within Central and Commercial Banking." Journal of Computer Technology and Applied Mathematics 1.1 (2024): 85-92.

[32] Qin, L., Zhong, Y., Wang, H., Cheng, Q., & Xu, J. (2024). Machine Learning-Driven Digital Identity Verification for Fraud Prevention in Digital Payment Technologies.
[33] Bai, Xinzhu, Wei Jiang, and Jiahao Xu. "Development Trends in AI-Based Financial Risk Monitoring Technologies." Journal of Economic Theory and Business Management 1.2 (2024): 58-63.
[34] Jiang, W., Yang, T., Li, A., Lin, Y., & Bai, X. (2024). The Application of Generative Artificial Intelligence in Virtual Financial Advisor and Capital Market Analysis. Academic Journal of Sociology and Management, 2(3), 40-46.
[35] Zhong, Y., Cheng, Q., Qin, L., Xu, J., & Wang, H. (2024). Hybrid Deep Learning for AI-Based Financial Time Series Prediction. Journal of Economic Theory and Business Management, 1(2), 27-35.
[36] Xiang, A., Qi, Z., Wang, H., Yang, Q., & Ma, D. (2024). A Multimodal Fusion Network For Student Emotion Recognition Based on Transformer and Tensor Product. arXiv preprint arXiv:2403.08511.
[37] Han, Wang, et al. "ROBO-ADVISORS: REVOLUTIONIZING WEALTH MANAGEMENT THROUGH THE INTEGRATION OF BIG DATA AND ARTIFICIAL INTELLIGENCE IN ALGORITHMIC TRADING STRATEGIES." Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online) 3.3 (2024): 33-45.
[38] Liang, P.; Song, B.; Zhan, X.; Chen, Z.; Yuan, J. Automating the training and deployment of models in MLOps by integrating systems with machine learning. Appl. Comput. Eng. 2024, 67, 1−7, https://doi.org/10.54254/2755-2721/67/20240690.
[39] Zhou, Y.; Zhan, T.; Wu, Y.; Song, B.; Shi, C. RNA secondary structure prediction using transformer-based deep learning models. Appl. Comput. Eng. 2024, 64, 95–101, https://doi.org/10.54254/2755-2721/64/20241362.
[40] Liu, B., Cai, G., Ling, Z., Qian, J., & Zhang, Q. Precise Positioning and Prediction System for Autonomous Driving Based on Generative Artificial Intelligence. （10个）
[41] Cui, Z., Lin, L., Zong, Y., Chen, Y., & Wang, S. Precision Gene Editing Using Deep Learning: A Case Study of the CRISPR-Cas9 Editor.
[42] Wang, B., He, Y., Shui, Z., Xin, Q., & Lei, H. (2024). Predictive Optimization of DDoS Attack Mitigation in Distributed Systems using Machine Learning. Applied and Computational Engineering, 64, 95-100.