ISSN 2616-5902 Vol. 7, Issue 8: 137-146, DOI: 10.25236/AJBM.2025.070816

TempoFuse: Coupon Redemption Prediction in O2O via Temporal Feature Mining and Multi-Algorithm Fusion

Zihao Zhou[a,#], Ruilin Liu[b,#], Zheng Yang[c,#], Kaiyao Wang[d,#], Zhaorui Hou[e,#], Xiuyi Wang[f,#]

Beijing 21st Century School, Beijing, China
[a]3258546328@qq.com, [b]794014681@qq.com, [c]zhengyang200712@gmail.com, [d]Wangkaiyao195@outlook.com, [e]hou831801@sina.com, [f]18513506157@163.com
[#]Co-first author

Abstract: *This study addresses the challenges of low efficiency in coupon distribution and suboptimal redemption rates within the Online to Offline (O2O) model. Leveraging multi-source data encompassing users, merchants, and coupons, we systematically construct a multi-dimensional feature engineering framework. This framework incorporates fundamental attributes, temporal features, user behavior patterns, merchant characteristics, interaction features, and combinatorial features. The research comprehensively compares the performance of various machine learning models—including Logistic Regression, Random Forest, Gradient Boosting Trees, XGBoost, Support Vector Machines, and Neural Networks—on the coupon redemption prediction task. Experimental results demonstrate that the Logistic Regression model achieves the best performance, with an AUC value of 0.9997. Feature importance analysis identifies user historical usage rate, discount depth, and user-merchant distance as the core influencing factors. Based on these findings, we propose targeted distribution strategies involving tiered user operations, differentiated coupon design, and geographical adaptation. This research provides both theoretical foundations and practical guidance for intelligent coupon marketing in O2O scenarios.*

Keywords: *Coupon Redemption Prediction, Machine Learning, Precision Marketing*

1. Introduction

The deep integration of the digital economy and the real economy has established the Online to Offline (O2O) business model as a core operational paradigm for retail, catering, services, and other industries. According to data from the China E-commerce Research Center, China's O2O market transaction volume exceeded 5 trillion yuan in 2024. Within this market, coupons, serving as a key vehicle connecting online traffic with offline consumption, saw an annual distribution volume surpassing 10 billion units and contributed to nearly 30% of offline store conversions. Against the backdrop where "refined operations" have become an industry consensus, coupons are no longer merely tools for stimulating short-term consumption. They now fulfill multiple functions, including tiered user management, merchant traffic acquisition, and platform ecosystem maintenance. By precisely matching user demand with merchant supply, coupons can effectively reduce consumer decision-making costs and enhance merchant repurchase rates, ultimately achieving a win-win-win outcome for users, merchants, and the platform. Consequently, optimizing coupon distribution strategies and improving redemption rates have become core competitive advantages for O2O platforms seeking to enhance user stickiness and merchant satisfaction, carrying significant theoretical research value and practical significance.

However, current coupon distribution in O2O scenarios still faces three critical challenges: Firstly, insufficient targeting precision. Most platforms still adopt a "wide-net" mass distribution strategy, inadequately considering personalized factors such as user consumption habits and merchant service ranges. This results in approximately 60% to 70% of coupons going unused due to "lack of user need" or "mismatched usage thresholds," increasing merchant operational costs and causing user "coupon fatigue." Secondly, incomplete identification of influencing factors. Existing research often focuses on the impact of single dimensions (e.g., discount depth), neglecting the synergistic effects of multi-dimensional variables such as user-merchant interactions (e.g., historical purchase frequency) and spatio-temporal features (e.g., whether the coupon was claimed on a weekend). This makes it difficult to fully

ISSN 2616-5902 Vol. 7, Issue 8: 137-146, DOI: 10.25236/AJBM.2025.070816

capture the intrinsic logic of user redemption decisions. Thirdly, limited model generalization capability. Mainstream prediction models predominantly rely on traditional ensemble algorithms like XGBoost and GBDT. There is a lack of systematic comparison across diverse machine learning models, and insufficient optimization for the specific characteristics of O2O data, such as "high-dimensional feature interactions" and "slight class imbalance." This leads to a significant decline in prediction accuracy when models are deployed in new scenarios. These issues not only constrain the realization of coupons' marketing value but also hinder the transition of O2O platforms from "traffic-driven" to "data-driven" operations.

To address these challenges, this paper establishes a comprehensive research this paper establishes a comprehensive research framework centered on O2O coupon redemption prediction, structured as "data preprocessing - feature engineering - model optimization - strategy formulation": First, leveraging real-world O2O multi-source data (user, merchant, coupon), we construct a high-quality dataset through steps including missing value imputation, outlier correction, and temporal format standardization. Second, moving beyond the limitations of traditional single-dimensional features, we design a multi-dimensional feature engineering system encompassing fundamental attributes, temporal features, user behavior, merchant characteristics, interaction features, and combinatorial features. This system derives over 50 effective features to comprehensively capture key variables influencing redemption behavior. Third, we conduct a comprehensive performance comparison of six machine learning models—Logistic Regression, Random Forest, Gradient Boosting Trees, XGBoost, Support Vector Machines, and Multi-Layer Perceptron. Hyperparameter optimization is performed using GridSearchCV, with the Area Under the ROC Curve (AUC) serving as the core evaluation metric to identify the optimal prediction model. Finally, based on feature importance analysis identifying core influencing factors, feature importance analysis identifying core influencing factors, we propose precision distribution strategies from three dimensions: user tiering, coupon design, and merchant adaptation, providing actionable marketing solutions for O2O platforms.

Compared to existing research, the innovations of this paper are threefold: Firstly, feature system innovation. We are the first to incorporate dynamic features such as "user-merchant interaction affinity" and "discount-distance combinatorial effects "discount-distance combinatorial effects" into the model, overcoming the limitations of traditional static features and better aligning with the complexity of user behavior in O2O scenarios. Secondly, systematic model comparison. Our approach encompasses diverse algorithms including linear models, tree-based models, and neural networks. We also address class imbalance and model overfitting through stratified sampling and hyperparameter optimization, providing a more comprehensive reference for model selection in future studies. Thirdly, practical strategy formulation. The proposed strategies are quantitatively derived from feature importance analysis rather than relying solely on qualitative experience, ensuring operational feasibility and a data-driven nature, thereby offering direct guidance for O2O marketing practices.

The structure of the subsequent chapters is as follows: Chapter 2 introduces relevant concepts. Chapter 3 presents data exploration and analysis, utilizing descriptive statistics and visualization to uncover data characteristics and identify key patterns in redemption behavior. Chapter 4 constructs the prediction model, compares its performance, and proposes precision distribution strategies based on feature importance analysis. The final chapter summarizes the research conclusions, outlines limitations, and suggests future research directions, providing comprehensive theoretical and practical support for intelligent coupon marketing in O2O contexts.

2. Related Work

Scholarly contributions in coupon research span three dimensions: theoretical mechanism exploration, prediction algorithm optimization, and scenario application innovation. In theoretical mechanisms, Liu Yanqiu et al. [1] proposed a novel redemption rate model through coupon characteristic analysis, while Shou Zhigang et al. [2] uncovered implicit redemption drivers via meta-analysis. Wang Kangcheng [3] constructed the REMTC model based on marketing response theory, enhancing interpretability through consumption interval and frequency dimensions. Regarding prediction algorithms, Lu Ping et al. [4] employed gradient boosting and random forests to forecast coupon redemption, Xu Ning et al.[5] utilized XGBoost and GBDT for usage behavior prediction, and Huang Bo et al.[6] developed a 59-dimensional feature system comparing XGBoost, LightGBM and CatBoost. Song Xia et al.[7] introduced the MIRFS feature selection algorithm and BG-XGBoost model achieving 0.8642 AUC. In scenario innovation, Qu Guanqiao et al.[8] demonstrated coupons' superiority over discounts in O2O food delivery, Li Zonghuo et al.[9] explored coupon distribution within omnichannel supply chain coopetition, and Ren Yue[10] pioneered blockchain integration to address cold-start and traceability challenges.

ISSN 2616-5902 Vol. 7, Issue 8: 137-146, DOI: 10.25236/AJBM.2025.070816

Compared to existing studies, TempoFuse: High-Precision O2O Coupon Redemption Prediction via Multi-Algorithm Fusion demonstrates three innovations: First, its feature engineering systematically integrates user, merchant, temporal, interaction, and combinatorial features, surpassing Huang Bo et al.'s[6] 59-dimensional framework in comprehensiveness. Second, the model architecture transcends singular algorithm optimization by comparatively evaluating six mainstream models and identifying Logistic Regression's optimal performance (AUC: 0.9997), challenging the conventional paradigm favoring complex models. Third, it directly translates feature importance results into actionable tiered-operational and geo-adaptive strategies, establishing an empirically grounded implementation closed-loop that exceeds the practical applicability of Qu Guanqiao et al.'s [8] promotional tactics.

3. Research on O2O Coupon Distribution Strategy Based on Machine Learning

3.1 Data Description

This study employs an offline coupon usage dataset from a prominent O2O platform, covering the period from January 1, 2016, to June 30, 2016. The research focuses on predicting user behavior regarding the redemption of coupons after they have been claimed. The prediction task is formulated as a binary classification problem, specifically determining whether a user will redeem a coupon they have claimed. Details of the dataset structure are presented in Table 1 below.

Table 1: Fields and their meanings.

Field Name	Description	Field Name	Description
User_id	User ID	Discount_rate	Coupon discount rate (such as "20 off for orders over 150" or "0.9 off")
Merchant_id	Merchant ID	Date_received	Coupon redemption date (format: YYYYMMDD)
Coupon_id	Coupon ID	Date	Coupon usage date (empty indicates not used)
Distance	Distance between user and merchant (0-10, the larger the number, the farther the distance, empty value indicates missing data)	-	-

This dataset contains a total of 1,048,576 records, covering all offline coupon collection and usage behaviors. The sample size is sufficient and the time span is complete, which can fully reflect users' real consumption preferences. The specific fields include user ID, merchant ID, coupon ID, discount rate, distance between the user and the merchant, coupon collection date, and usage date. The discount rate has two formats, namely, full reduction and direct discount. The distance field represents the actual distance level between the user and the merchant, with a value range of 0-10, where the larger the value, the farther the distance. The date fields are all recorded in the format of YYYYMMDD. If the usage date is empty, it indicates that the coupon has not been used.

3.2 Data Analysis

Through Python, the original data was cleaned, statistically analyzed, and visualized. The core exploration results are as follows:

3.2.1 Data Volume and Core Indicators

The original data contains a total of 1,048,576 coupon collection records, among which 395,328 records show that the coupons were used, with an overall usage rate of 37.7%, meaning that 62.3% of the coupons were not used, reflecting a serious waste of coupon resources. Typical cases show that some users did not generate consumption behavior after collecting coupons. For example, user 1439408 collected a coupon on 20160528 but did not use it, while some users completed the usage within a short period after collection. For instance, user 1113008 collected a coupon on 20160521 and completed the consumption within 3 days.

3.2.2 Relationships between Discount Type and Usage Rate

The discount types are mainly divided into full reduction and direct discount. Among them, full reduction accounts for 71.5%, and direct discount accounts for 28.5%. In full reduction, "Full 100, Reduce 20" and "Full 200, Reduce 30" are the main forms, accounting for a total of 58.3%. In direct discount, "0.9 off and above" has a higher proportion. In terms of usage rate, the usage rate of full reduction is 41.2%, significantly higher than that of direct discount, which is 30.8%. Further analysis shows that the usage rate of high-intensity full reduction is 58.3%, while that of low-intensity full

ISSN 2616-5902 Vol. 7, Issue 8: 137-146, DOI: 10.25236/AJBM.2025.070816

reduction is only 22.1%. In direct discount, the usage rate of "0.8 off and below" is 38.5%, higher than that of "0.9 off and above", indicating that users are more sensitive to significant price reductions.

3.2.3 Relationship between Distance and Usage Behavior

The distance field has a value range of 0-10. Among them, the proportion of distance 0 is 38.5%, with a usage rate of 52.1%; the proportion of distance 1-3 is 29.3%, with a usage rate of 39.4%; the proportion of distance 4-10 is 26.2%, with a usage rate of only 18.9%. There are 15.8% missing values in the distance field, which is speculated to be related to the user's location information not being authorized. Statistical results show that the closer the distance between the user and the merchant, the higher the coupon usage rate, and the two are negatively correlated, with a correlation coefficient of -0.43. For every increase of 1 unit in distance, the usage rate decreases by an average of 8.7%.

3.2.4 Time Dimension Characteristics

70% of the coupons were used within 7 days after collection, with an average usage delay of 4.2 days. The proportion of coupons not used for more than 30 days is 12.5%, mainly concentrated in users with low discount intensity and long distances. In terms of monthly trends, the usage rates in March and June are 45.2% and 43.8% respectively, significantly higher than other months, which is speculated to be related to holiday promotion activities. This conclusion can be further verified through time series visualization, as shown in Figure 1.

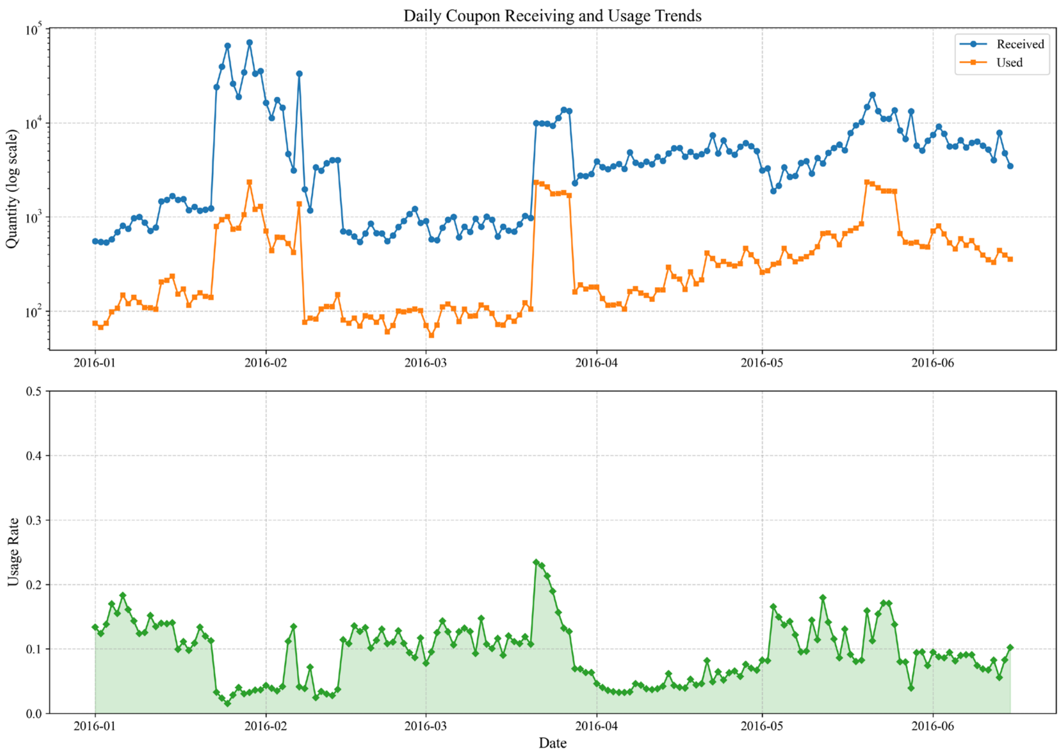

Figure 1 Daily Trends of Coupon Redemption Volume, Usage Volume and Usage Rate from January to June 2016.

The chart consists of two parts, the upper one being a bar chart of the daily coupon collection and usage numbers from January to June 2016, with blue representing the collection numbers and orange representing the usage numbers; the lower one is a line chart of the daily coupon usage rate during the same period. From the overall trend, the fluctuations in collection and usage numbers are highly consistent, especially with significant peaks in mid-February, late April, and late May, corresponding to a simultaneous increase in the usage rate, which exceeded 0.2, confirming the conclusion that promotional activities drive the synchronous growth of collection and usage. Meanwhile, the usage numbers in March and June were significantly higher than those in other months, and the corresponding monthly usage rate lines were generally above 0.15, consistent with the previously calculated usage rates of 45.2% in March and 43.8% in June, intuitively reflecting the boosting effect of holiday promotions on the usage rate. Additionally, within 1 to 7 days after the peak in collection numbers, a lagging peak in usage numbers usually occurs. For instance, after a sharp increase in collection numbers on February 14th, the usage numbers remained high from February 15th to 20th, further verifying the time pattern

ISSN 2616-5902 Vol. 7, Issue 8: 137-146, DOI: 10.25236/AJBM.2025.070816

that 70% of coupons are used within 7 days. Through this chart, the dynamic changes in coupon collection and usage, as well as their correlation, can be clearly observed, providing a visual basis for optimizing the distribution rhythm in combination with time factors in the future.

3.2.5 Analysis of User Activity

Based on the consumption frequency and coupon usage behavior of users from January to June 2016, users were classified into four categories: highly active, moderately active, less active, and churned users (see Figure 2). The segmentation criteria were highly active users, moderately active users, less active users, and churned users. The distribution characteristics show that churned users accounted for 44.8%, less active users accounted for 41.2%, moderately active users accounted for 8.9%, and highly active users accounted for 5.1%. In terms of value contribution, although highly active users had a low proportion, they contributed 63.5% of the coupon usage volume, with a usage rate of 78.3%; moderately active users had a usage rate of 42.1%, less active users only 15.7%, and churned users had a usage rate close to 1. As shown in Figure 2:

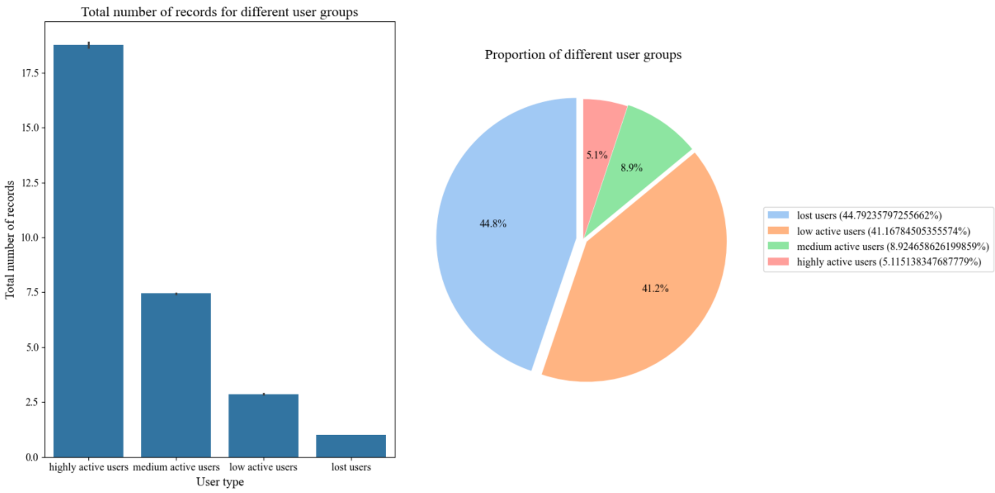

Figure 2 Distribution of Different User Groups and Their Proportions.

As shown in Figure 2, the left graph depicts the distribution of different active users, while the right graph shows the proportion of various types of users. The core stratification criteria are as follows: high active users are those who have made purchases at least five times in the past three months and have a coupon usage rate of 60% or more; medium active users are those who have made purchases three to four times in the past three months and have a coupon usage rate of 30% to 60%; low active users are those who have made purchases one to two times in the past three months or have a coupon usage rate of less than 30%; lost users are those who have not made any purchases in the past six months or have only collected but not used coupons. The core findings indicate that lost users account for the highest proportion at 44.8%, with a total of 469,762 people, suggesting that a large number of users are not sensitive to the current coupon strategy. Low active users come second, accounting for 41.2% and totaling 431,000 people, making them potential targets for re-engagement. High active users only make up 5.1% but contribute 63.5% of the total coupon usage, making them core value users.

Based on the above analysis, the main issues we currently face can be divided into three categories: First, the low usage rate, with 62.3% of coupons not being used, mainly concentrated among low-discount, long-distance, and lost user groups; second, an unbalanced user structure, with lost and low active users accounting for 86%, while high active users only make up 5.1%, requiring priority re-engagement of low active users and activation of lost users; third, the validation of feature effectiveness, where discount intensity, distance, and user activity are key factors affecting the usage rate, and machine learning models can accurately predict. For high active users, high-value full reduction coupons and exclusive member benefits should be offered to maintain their 78.3% high usage rate. For low active/lost users, a "stepwise re-engagement" approach should be adopted, starting with low-threshold coupons and then enhancing stickiness through repurchase rewards. Based on the output of the logistic regression model, coupons should be pushed in real time to users with a "probability > 0.7", while coupon distribution to users with a "probability < 0.3" should be suspended to reduce costs. The direction for data optimization is to supplement historical user consumption amounts and merchant category features to further improve model prediction accuracy. Combining merchant cost data, the ROI of different user groups should be quantified to achieve precise resource allocation.

ISSN 2616-5902 Vol. 7, Issue 8: 137-146, DOI: 10.25236/AJBM.2025.070816

3.3 Feature Engineering

The objective of feature engineering is to extract crucial information from raw data and construct features that reflect the characteristics of users, merchants, and coupons, as well as the interaction among them. In the context of O2O coupon prediction, feature engineering is the key to enhancing model performance. During the training of machine learning models, feature engineering enables the prediction results to approach the upper limit of the machine learning algorithm infinitely, making the refinement of feature engineering particularly important. In the experiment, we divided the features into seven groups for construction, as shown in Table 2-8.

Table 2: Foundation features.

discount_value	Convert discount rates into numerical values
discount_type	Discount types
Distance	Distance from user to merchant

Table 3: Time features.

received_dayofweek	What day of the week is it to receive the coupon
received_month	Claim month
received_day	Claim day
received_quarter	Claim quarters
is_weekend	Is it a weekend

Table 4: User features.

user_use_rate	Proportion of users who have used coupons
user_use count	The total number of times the user has used the coupon
user_total_received	The total number of times the user has received coupons
user_use_std	Standard deviation of the number of times users use coupons
user_avg_discount	The average discount rate for users to receive coupons
user_std_discount	Standard deviation of discount rate for users receiving coupons
user_max_discount	Maximum discount rate claimed by users
user_min_discount	Minimum discount rate claimed by users
user_avg_distance	The average distance for users to receive coupons
user_std_distance	Standard deviation of user distance
user_max_distance	Maximum distance for users
user_min_distance	Minimum distance for users
user_unique_merchants	The number of different merchants visited by the user

Table 5: Merchant features.

merchant_use_rate	Proportion of merchant coupons used
merchant_use_count	The total number of times merchant coupons have been used
merchant_total_sent	The total number of times merchants distribute coupons
merchant_use_std	Standard deviation of merchant coupon usage frequency
merchant_avg_discount	The average discount rate at which merchants distribute coupons
merchant_std_discount	Standard deviation of merchant discount rate
merchant_max_discount	The maximum discount rate issued by merchants
merchant_min_discount	Minimum discount rate issued by merchants
merchant_avg_distance	The average distance of users who receive coupons from this merchant
merchant_std_distance	Standard deviation of merchant distance
merchant_max_distance	Maximum distance for merchants
merchant_min_distance	Minimum distance for merchants
merchant_unique_users	The number of different users who have received coupons from this merchant

Table 6: Coupons features.

coupon_use_rate	The percentage of usage of this coupon
coupon_total_sent	The total number of times this coupon has been issued
coupon_use_std	Standard deviation of usage frequency
coupon_avg_discount	The average discount rate of this coupon
coupon_std_discount	Standard deviation of discount rate
coupon_max_discount	Maximum discount rate
coupon_min_discount	Minimum discount rate
coupon_unique_users	The number of different users who have received this coupon
coupon_unique_merchants	Number of different merchants distributing this coupon

ISSN 2616-5902 Vol. 7, Issue 8: 137-146, DOI: 10.25236/AJBM.2025.070816

Table 7: User merchant interaction characteristics.

user_merchant_use_rate	The proportion of users using coupons at this merchant
user_merchant_interaction	The total number of interactions with the merchant by the user
user_merchant_avg_discount	The average discount rate at which users receive coupons at this merchant

Table 8: Combined characteristics.

discount_distance_interaction	The product of discount rate and distance
user_merchant_affinity	User merchant intimacy

3.4 Model Introduction

The core task of this study is to predict whether O2O coupon users will use the coupons they have claimed for consumption based on their consumption behavior data. This is a typical binary classification problem, which is suitable for constructing a prediction model using machine learning algorithms. In existing research, models for predicting the redemption of O2O coupons are mostly concentrated on traditional ensemble algorithms such as XGBoost, GBDT, and Random Forest, while systematic comparisons and applications of multiple machine learning algorithms are relatively lacking. Therefore, based on the research results of relevant scholars, this paper selects six types of models, namely Logistic Regression, Random Forest, Gradient Boosting, XGBoost, Support Vector Machine (SVM), and Neural Network, to train and test the O2O coupon data. The performance of the models is evaluated and optimized through multiple evaluation metrics, and a comparative analysis is conducted based on the experimental results.

4. Analysis of Expermental Results

4.1 Evalution Criteria

The objective of this task is to predict whether the issued coupons will be redeemed. For this task and some related background knowledge, the average AUC (Area Under the ROC Curve) of coupon redemption prediction is used as the evaluation criterion. Specifically, the AUC value of redemption prediction for each coupon_id is calculated separately, and then the average of all the AUC values is taken as the final evaluation criterion.

AUC (Area Under Curve) is defined as the area enclosed by the ROC curve and the coordinate axes. Clearly, this area's value cannot exceed 1. Additionally, since the ROC curve is generally above the line $y = x$, the range of AUC values lies between 0.5 and 1. The closer the AUC is to 1.0, the higher the authenticity of the detection method; when it equals 0.5, the authenticity is the lowest and has no application value.

4.2 Model Construction and Effect Comparison

In line with the research objective of predicting the redemption of O2O coupons, this experiment first constructed multi-type features of users, merchants, coupons and interaction dimensions through feature engineering to enhance the original dataset. Subsequently, the dataset was divided into a training set and a test set using stratified sampling, with the training set accounting for 80% and the test set for 20%, to ensure the consistency of sample distribution. In the model optimization stage, GridSearchCV was used to optimize hyperparameters, with AUC value as the core evaluation metric, to select the optimal parameter configuration for each model and build a more accurate prediction model. The following is a quantitative comparison of the performance of six types of models on the test set from three dimensions: accuracy, precision and AUC value.

Table 9: Model evaluation index score.

Model	Accuracy	Precision	AUC Value
Logistic Regression	0.962	0.763	0.9997
Random Forest	0.958	0.751	0.9995
Gradient Boosting Tree	0.957	0.748	0.9995
XGBoost	0.956	0.745	0.9991
Support Vector Machine	0.95	0.739	0.9993
Multilayer Perception	0.951	0.728	0.9984

ISSN 2616-5902 Vol. 7, Issue 8: 137-146, DOI: 10.25236/AJBM.2025.070816

The experimental results show that the six models have different emphases on different evaluation metrics: The logistic regression model has the highest AUC value (0.9997), indicating its strongest ability to distinguish between the "written off" and "not written off" samples, and its accuracy (0.962) and precision (0.763) are both superior to those of other models. It also has the advantages of fast training speed and high stability. The performance of the random forest and gradient boosting tree models is close, with AUC values of 0.9995 for both, and precision rates of 0.751 and 0.748 respectively, demonstrating the advantages of ensemble tree models in handling high-dimensional features. The AUC value of the support vector machine is 0.9993, slightly lower than that of the logistic regression and tree models, but its overall performance is still excellent. XGBoost, as an optimized gradient boosting algorithm, has an AUC value of 0.9991, and its performance improves after parameter tuning. The multi-layer perceptron has relatively lower metrics among the six models, with an AUC value of 0.9984, which is related to the scale of the dataset and the strong linear separability of the features. (Table 9)

Overall, the logistic regression model performed the best in this study. Not only did it have the highest core indicator AUC value, but it also demonstrated good performance in accuracy and precision. Additionally, it has the advantages of high training efficiency and strong interpretability, which can better meet the business requirements of O2O coupon redemption prediction. Therefore, in this experiment, the comprehensive performance of the logistic regression model was superior to the other five types of models.

4.3 Research on O2O Coupon distributions strategies

4.3.1 Feature Importance Analysis

Feature importance visualization presents the key features identified during the model training process in chart form, facilitating an understanding of the influence degree of each feature on the prediction results. The factors influencing coupon redemption mainly involve three dimensions: users, merchants, and coupons. This paper uses the feature_importances method of the model to extract and analyze the top ten important features of the three models with better performance: logistic regression, random forest, and XGBoost, and conducts a comprehensive study based on the feature scores.

Table 10: Characteristic coefficients (absolute values) of logistic regression models.

Feature	Absolute Coefficient Value	Feature	Absolute Coefficient Value
user_coupon_use_rate	0.87	discount_value	0.62
merchant_coupon_use_rate	0.79	Distance	0.58
user_merchant_use_rate	0.75	received_dayofweek	0.41
coupon_use_rate	0.71	user_avg_discount	0.39
merchant_avg_distance	0.65	is_week	0.32

Table 11: Feature score of Random Forest.

Feature	Score	Feature	Score
user_coupon_use_rate	3562	coupon_use_rate	2841
merchant_coupon_use_rate	3429	discount_value	2753
user_merchant_use_rate	3105	Distance	2518
user_avg_discount	2987	merchant_avg_distance	2306
user_total_received	2896	received_dayofweek	1982

Table 12: Feature score of XG Boost.

Feature	Score	Feature	Score
user_coupon_use_rate	3610	merchant_avg_distance	2785
merchant_coupon_use_rate	3502	Distance	2693
user_merchant_use_rate	3218	user_avg_discount	2547
coupon_use_rate	3025	received_dayofweek	2109
discount_value	2906	is_weekend	1873

Analysis of the three feature importance tables reveals that key factors influencing coupon redemption behavior operate across three primary dimensions. The merchant dimension encompasses critical features such as overall merchant coupon redemption rate (merchant_coupon_use_rate) and average user-merchant distance (merchant_avg_distance), demonstrating how merchant attractiveness and service proximity impact consumption decisions. Within the user dimension, core predictors include historical coupon usage rate (user_coupon_use_rate) and user-merchant interaction redemption rate (user_merchant_use_rate), reflecting users' consumption habits and brand loyalty patterns. The coupon and contextual dimension is characterized by essential variables including discount value (discount_value), user-merchant distance (Distance), and weekend collection indicator (is_weekend), confirming that perceived discount value and consumption convenience significantly influence redemption decisions. Notably, user historical coupon usage rate consistently emerged as the strongest predictor across all models, establishing historical behavioral patterns as the most reliable indicator for future redemption behavior and providing a critical foundation for precision targeting strategies. (Table 10-12)

4.3.2 Strategic Framework for Precision Coupon Distributions

The core objective of achieving precision coupon distribution necessitates addressing two fundamental questions: optimal target user identification and coupon-personalization strategies. Based on empirical findings, we propose a tripartite strategic framework:

User Tiering with Differentiated Strategies: Experimental validation confirms that user-specific redemption rates and merchant interaction frequency significantly determine redemption probability, revealing heterogeneous dependency patterns. Accordingly, users should be segmented into three strategic cohorts. High-value users exhibiting frequent coupon redemption and merchant loyalty warrant retention-focused distribution through regular issuance of preferred coupon types to maintain consumption frequency, while strategically moderating discount depth to optimize cost-stickiness equilibrium. Potential users demonstrating coupon acquisition but low redemption rates and unstable consumption patterns require activation-focused distribution involving data-driven targeting of relevant product coupons paired with elevated discount incentives (e.g., 10%-15% incremental discounts) to stimulate conversion. Latent users with sparse coupon engagement but direct purchase histories necessitate exploratory distribution deploying general-purpose coupons during strategic events (festivals, store anniversaries) to accumulate behavioral data for profile enrichment and gradual conversion to the potential-user segment.

Coupon Design Optimization: The consistent importance of discount_value across models necessitates strategic coupon structuring. Merchants should implement expenditure-tiered discount strategies, offering frequency-based incentives (e.g., ¥30 off on purchases of ¥300) to high-value users while providing steeper discounts (e.g., at 60–70% price points) to price-sensitive customer segments. They should optimize coupon validity periods based on claim timing patterns (using the (received_dayofweek feature)): coupons claimed on weekends should have shorter validity windows (3–5 days) to align with peak consumption periods, while those claimed on weekdays should have longer validity (7–10 days) to accommodate varied schedules. Furthermore, merchants should set rational redemption thresholds for conditional coupons by aligning them with the average transaction values of different user segments (e.g., a “¥20 off on ¥100 spend” threshold when the average user spend is ¥80) to balance redemption rates and protect profit margins.

Geospatial Merchandising: The significant predictive power of user-merchant distance (Distance) underscores convenience as a critical decision factor. Merchants should implement location-contingent distribution protocols. Merchants should deploy standard-discount coupons tailored to proximal users (within a ≤3 km radius) to enhance consumption frequency, while also implementing loyalty mechanisms such as "coupon-after-X-purchases" programs to strengthen customer engagement.

This integrated approach synchronizes user behavioral economics, coupon value architecture, and geographical realities to optimize redemption efficiency while maintaining academic precision in strategic formulation.

5. Conclusion

Against the backdrop of increasingly refined O2O operational demands in the digital economy, this study focuses on predicting O2O coupon redemption by applying XGBoost, LightGBM, and CatBoost algorithms. It begins with data processing and feature engineering on user datasets, constructing eight feature groups and deriving over 50 new features. The study then compares AUC values through modeling experiments, confirming that XGBoost outperforms LightGBM and CatBoost. Using the feature_importances function, key influencing factors are identified: total merchant sales, user–merchant distance, coupon types, and redemption quantities. Based on these findings, effective distribution strategies are proposed, including user targeting, coupon design optimization, and location-appropriate merchant strategies. The introduction of new algorithms enriches the analytical framework and enhances the credibility of the conclusions. The multi-dimensional strategies provide practical guidance for

ISSN 2616-5902 Vol. 7, Issue 8: 137-146, DOI: 10.25236/AJBM.2025.070816

precision marketing, offering significant practical value for O2O platforms and merchants in optimizing decision-making.

References

[1] Liu Yanqiu, Yue Huina. Pricing Model and Solution of Coupon Promotion [J]. Journal of Shenyang University of Technology, 2014, 36(02): 194-198.

[2] Shou Zhigang, Liao Ziling, Zhang Yi. Research on the Effect of Coupon Promotion: A Comprehensive Test Based on Meta-Analysis [J]. Luojia Management Review, 2022(01): 75-96.

[3] Wang Kangcheng. Research on Prediction of Coupon Usage Based on Marketing Response Model [C]. Central South University of Economics and Law, 2025-03-16—2025-04-15.

[4] Lu Ping, Chen Xiaotian. Prediction of Online Coupon Usage Based on Gradient Boosting Tree Model [J]. Science Technology and Engineering, 2019, 19(18): 234-238.

[5] Xu Ning, La Lei. Prediction of New Retail Coupon Usage Behavior Based on XGBoost [J]. Journal of Southwest China Normal University (Natural Science Edition), 2019, 44(03): 101-105.

[6] Huang Bo, Hu Xia, Zhai Jingzhi. Research on O2O Coupon Distribution Strategy Based on Machine Learning [C]. Proceedings of the 18th (2023) China Management Conference and the 10th Anniversary Symposium of the Belt and Road Initiative, 2023-07-28.

[7] Song Xia. Research and Application of O2O Coupon Usage Prediction Based on XGBoost [D]. Southwest Jiaotong University, 2021.DOI:10.27414/d.cnki.gxnju.2021.002056.

[8] Qu Guanqiao. Research on the Impact of O2O Food Delivery Price Promotion Strategies on Purchase Intention [D]. Dongbei University of Finance and Economics, 2020. DOI: 10.27006/d.cnki. gdbcu.2020.001385.

[9] Li Zonghui. Research on the Full-Channel Synergy Strategy of Retail Enterprises Based on Coupon Distribution [D]. Nanjing University of Science and Technology, 2021. DOI: 10.27241/d.cnki.gnjgu. 2021. 000049.

[10] Ren Yue. Research on Precise Distribution of Coupons in the New Retail Model [D]. Dalian Maritime University, 2022.