Implementation Path of Machine Learning in Intelligent Marketing Scenarios of Commercial Banks

Chunliang Yang [1, a], and Jiuzhou Cheng [2, b]

[1]Shandong Sundata Data Management Software Engineering Co.Ltd, Shandong,China;

[2] Shandong Experimental High School, Shandong,China.

[a] yang_cliang@sundatasoft.com, [b] 656362656@qq.com

Abstract. This paper takes commercial banks' intelligent marketing scenario as the entry point and introduces machine learning into the intelligent marketing scenario. It introduces the range of application and implementation path of machine learning in commercial banks' intelligent marketing solutions, discusses the possible problems and methods in the process of model construction, and gives empirical analysis of simulating scenarios. Emphasis is placed on developing effective marketing strategies and adjusting marketing strategies based on the results of model deployment monitoring to establish a complete intelligent marketing system and closed-loop process. It provides reference for commercial banks to improve marketing response rate and realize digital operation.

Keywords: machine learning; intelligent marketing; marketing strategy; digital operations.

1. Intelligent Marketing Background

In recent years, the business environment of the traditional banking industry is undergoing profound changes. New technologies and new business models such as big data, cloud computing and mobile payment are emerging, and more and more bank managers realize that exploring and creating digital financial services and operation systems and promoting operation management and service methods are the keys to achieving strategic transformation and operational changes in banks.

While commercial banks are facing severe challenges and impacts, there are also full of opportunities. First of all, commercial banks have accumulated rich data after nearly decades of information technology construction; the IT foundation has been improved and the data governance capability has been gradually improved. Commercial banks are using big data to drive business and make operations more accurate and efficient.

Secondly, retail business is a must-compete area in the banking industry, and its development has become a sharp weapon for banks competing to seize the market. Major banks are accelerating digital transformation, strengthening online operation capabilities, innovating service models and activity, enriching scenario ecology, and providing customized and specialized financial products and services for customers. Therefore, accurate insight into customer needs, good customer classification and improved customer experience have become issues that banks must face at this stage.

2. Application of machine learning in intelligent marketing scenarios

Commercial banks have a wide variety of retail businesses and a large customer base. Under the general trend of technology-driven finance, the external environment and competition faced by retail business have been revolutionized. Relying on traditional marketing experience cannot reflect the value of data, and it is difficult to find precise target customers and meet existing marketing demands. Precision marketing based on machine learning is rapidly emerging in the domestic banking industry.

Aiming at the characteristics of machine learning and marketing scenarios, we introduce the range of application of machine learning in intelligent marketing scenarios from the whole life cycle of marketing.

In short, in the whole life cycle of customers, we create a intelligent marketing paradigm through machine learning, strengthen customer-centered information sharing and value creation, enhance the ability of customer group stratification management, promote intelligent and refined customer relationship management, and enhance customer stickiness and stability.

Intelligent marketing model construction implementation path

Taking the marketing of gold accumulation products as an example, we introduce the implementation path of machine learning in marketing scenarios and the possible problems and solutions in the modeling process.

Products:Gold Accumulation,Customers buy gold from the Bank by way of fixed deposit or active accumulation and deposit it into their Gold Accumulation Account.And the bank deducts the corresponding gold balance from their gold accumulation account.

Target group: customers who already hold financial products but have not yet joined the gold accumulation program in a commercial bank.

Business Definition explanation:Before building the model, the business language needs to be parsed into the model language, i.e., the following definitions need to be clarified: observation point, observation period, performance period, features, and target variables.

Data collection and integration:the first step is to collect and integrate data. Collecting means collecting user information and behavioral data in each channel. Integration means to identify the identity of users in each channel and to carry out unification.

Exploratory data analysis (EDA): it is a data analysis method that explores the structure and patterns of the data by graphing, tabulating, and calculating statistics on the original features to gain a prospective understanding of the original data.

Sample selection: The existing data set is usually divided into a training set and a test set, where the training set is used to train the model, and the test set is used to evaluate the discriminative ability of the model for new samples.

Feature engineering: it refers to the engineering approach to filter out better data features from the raw data and reduce the interference of noise to improve the training effect of the model. Feature engineering includes data cleaning, data preprocessing , feature selection, feature dimensionality reduction, etc.

Modeling and Tuning Parameters:the tuning of the parameters is time-consuming and laborious depending on the quality of the data. The goal is to achieve a perfect balance of variance and bias! In practice, we perform an initial range search first, and then narrow the range for a more refined search where good results appear, and get almost the same results before tuning the other parameters.

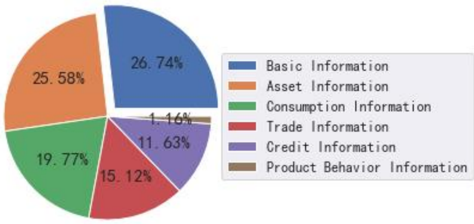
Model evaluation metrics are numerical values that measure the effectiveness of the model and are an assessment of the overall performance effectiveness of the model. Common evaluation metrics for classification models include confusion matrix, accuracy, precision, recall, etc.

Post-launch monitoring and tuning:After the model is developed, it does not mean the end of the project. After the model is trained, it needs to be improved according to the actual A/B testing and suggestions from business people, andtime partition testing.If the performance meets the expectation, the model could be launched.

3. Empirical analysis of simulation scenarios

In this chapter, we use the simulation data as an example and use the Python language to complete the empirical analysis of the construction of the Gold Accumulation intelligent marketing model in the previous chapter.

The flag in the simulation data source is the dichotomous tag 0/1 identification, custid is the customer code, JA001, JA002 ......JF015 are the names of the wide table feature variables after data preprocessing, where JA class feature indicates basic customer information, JB class feature indicates customer asset information, JC class feature indicates customer consumption information, JD class feature indicates customer transaction information, JE class feature indicates customer credit information, and JF class feature indicates product behavior information. Total 3636 records, 6 categories and 86 features.



Figur 1 : Distribution of the number of each type of features

After the missing value rate ranking analysis of the features, the features with missing rate > 0.25 were removed.

Table 1 Ranking of the missing rate of some features

Varname	count	rate
JA034	1369	0.376513
......	......	......

Descriptive statistical analysis was done on the remaining features to understand the approximate distribution of the sample features, with the following example results:

Table 2 Results of descriptive statistical analysis of selected features

	count	mean	std	min	25%	50%	75%	max
JA003	3636	5.67	234.50	-31.48	0.04	0.17	0.33	10000.00
......	......	......	......	......	......	......	......	......

The target variables were statistically analyzed as follows, where: 1 is a positive sample, representing customers who did not purchase before the observation point and purchased gold accumulation product during the performance period. 0 is a negative sample, representing customers who did not purchase before the observation point and still did not purchase gold accumulation product during the performance period. The distribution of positive and negative samples is about 1:22.

After the initial analysis and cleaning of samples and features as described above, the training and test sets are divided in the ratio of 7:3. The training set samples are woe-binned and the advantage of using woe binning coding is that the missing value processing and coding can be performed simultaneously.

Table 3: Results of woe binning for some characteristic variables

Varname	bin	ll	ul	total	woe	iv0	IV
JA003	1	-inf	-0.15	141.00	-1.14	0.04	0.097
	2	-0.15	-0.06	127.00	-1.04	0.03	
	3	-0.06	0.04	368.00	-0.21	0.01	
	4	0.04	inf	1909.00	0.13	0.01	
......	......	......	......	......	......	......	......

Feature selection can be performed based on the derived indicators IV values after woe binning. Here, IV<0.1 is used as the feature selection threshold, and 19 features with IV<0.1 after woe binning are deleted. The remaining features are analyzed for the binning results, and the feature variables whose monotonicity does not match the business meaning are deleted. The visualization of some features woe binning results are shown as follows:

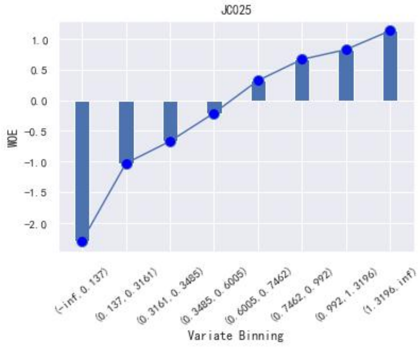


Figure 2 : woe binning results for feature JC025

The remaining features are woe coded on the training and test sets according to the binning rules to form the woe-binned replaced samples for later feature selection and model training.

The Pearson correlation coefficients and heat maps were calculated for the replaced training samples, and the two indicators with correlation coefficients greater than 0.5 were retained, and the indicators with relatively lower differentiation ability or stability were deleted. The calculation results and the heat map of correlation coefficients of some features of distinguishing ability and stability indicators are as follows:

Table 4: Calculation results of differentiation ability and stability index of some characteristic variables

Varname	IV	KS	AR	PSI
JA032	0.586345	0.244988	0.339399	0.009431
......	......	......	......	......

Next, the samples after feature engineering are trained with multiple alternative models, and the optimal model is determined by observing the performance of evaluation metrics such as AUC and accuracy on the test set.

Table5 : Performance of evaluation metrics of alternative models on the test set

Alternative Models	AUC	accuracy	precision	recall	F1	F2
svm	0.698	0.956	0.444	0.085	0.143	0.102
knn	0.706	0.957	0.5	0.106	0.175	0.126
logistic	0.798	0.741	0.117	0.766	0.203	0.363
mlp-nn	0.732	0.95	0.318	0.149	0.203	0.167
gbm	0.812	0.956	0.462	0.128	0.2	0.149
randomforest	0.788	0.952	0.222	0.043	0.071	0.051
naivebayes	0.742	0.63	0.08	0.723	0.144	0.277
xgboost	0.799	0.957	0.5	0.106	0.175	0.126

Considering the evaluation index results of each model, the logistic regression model was selected as the primary model. The results of partial evaluation indexes of the model training set and test set are as follows:

Table 6 Comparison of evaluation metrics between the training and test sets of the logistic regression model

Sample set	AUC	precision	recall	accuracy	F1	F2
Training set	0.861	0.127	0.827	0.746	0.22	0.393
Test set	0.798	0.114	0.766	0.732	0.198	0.356

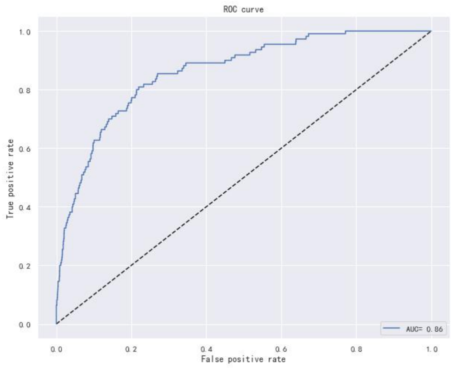


Figure 3 : ROC plot for the training set of the logistic regression model

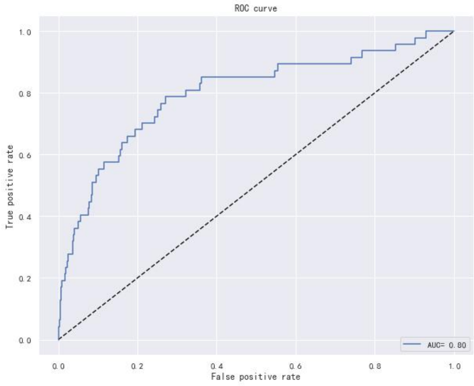


Figure 4 : Logistic regression model test set ROC curve

4. Intelligent Marketing Strategy Development

After constructing a supervised machine learning model based on the customer data of historical product purchases and generating a ranked list of customer purchase probabilities, it is necessary to combine the marketing model results to select the appropriate channels for different scenarios and conduct cross-channel marketing to implement the strategy.

When the strategy is implemented, each strategy should have specialized effect evaluation metrics, and the strategy should be monitored by data. Compare the conversion effect of different strategies in different stages, and iteratively optimize the operation strategies based on the data obtained from monitoring.

Scenario extension has become a new touch point in the Internet finance era. Commercial banks should try to integrate retail products into more scenarios, further increase the degree of penetration into the life scenarios of their customers, continuously unlock new financial service scenarios, and expand the relevance and coverage of scenarios.

Reference

[1] Liu Lixin, Liu Kaiqiang, Qiao Guiming. Research on methods and strategies of big data precision marketing for retail customers of commercial banks[J]. Rural Finance Research,2019(10):54-62.

[2] Wei, Bing-Quan, Yuan, Weidong, Qin, Si-Jia. Intelligent marketing model for inclusive credit customers based on data mining[J]. Financial Technology Times,2020,28(04):12-19.

[3] Pang Qiaohong. Application of data mining technology in credit card customer analysis[D]. Southwest University of Finance and Economics, 2012.

[4] Ruan Mengjiao,Gai Yunhao,Hu Xuyong,Liao Bo. Analysis of marketing strategy of personal financial products of commercial banks[J]. Modern Business,2021(05):139-141

[5] Pan Bo. Research on retail customer churn of Z bank based on data mining [D]. Zhejiang University of Technology,2019