Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.24499

Machine Learning Algorithm Based Ad Click Prediction and Marketing Competitive Analysis

Siqi Huang[1], Chenglin Ma[1*]

[1]*School of Finance and Economics, Hubei Engineering University New Technology College, XiaoGan, China*

[*]*Corresponding Author. Email: 705084526@qq.com*

Abstract. In this paper, by systematically evaluating the performance of multiple machine learning models in the task of advertisement click prediction, it is found that the XGBoost algorithm exhibits the best prediction potential by virtue of its integrated learning advantages. In order to further improve the model performance, the Sparrow Search Algorithm (SSA) is innovatively introduced to intelligently search for the key hyperparameters of XGBoost, and the SSA-XGBoost fusion model is constructed. The experimental results show that the optimized model achieves significant breakthroughs in classification performance: the accuracy rate reaches 0.87, which is 18.1% higher than that of the basic XGBoost; the recall rate is synchronously increased to 0.87, while the precision rate achieves a leapfrog growth to reach the excellent level of 0.887, which is 21.3% higher than that of the unoptimized model (0.731). These performance improvements have special value in the dimension of false alarm rate reduction - when the model accuracy rate is increased by 21.3%, it means that about 50,000 invalid placements can be reduced in a million-volume ad exposure scenario, and this accuracy improvement not only verifies the effectiveness of the sparrow search algorithm in parameter optimization, but also highlights the practical business value brought by the algorithm improvement. This improvement in accuracy not only verifies the effectiveness of the algorithm in terms of parameter optimization, but also highlights the practical commercial value of the algorithm improvement. From the perspective of feature engineering, SSA successfully solves the efficiency bottleneck of traditional grid search in high-dimensional parameter space through the strategy of combining global search and local optimization, so that key hyperparameters such as the tree structure parameters and learning rate of XGBoost reach a more optimal configuration, which effectively mitigates the risk of overfitting while maintaining the model's stronger generalization ability (19.2% improvement in F1-score) (34% reduction in cross-validation variance). The intelligent prediction model constructed in this study is of great practical significance to the field of digital marketing: through high-precision click prediction, advertisers can accurately identify potential user groups, and reduce the cost of ineffective advertisement exposure while improving the conversion efficiency. This data-driven decision support can not only optimize the advertising budget allocation strategy, but also promote the programmatic advertising delivery system to evolve in the direction of intelligence, and provide technical support for enterprises to build core advantages in digital marketing competition.

Keywords: Machine learning, Ad click prediction, XGBoost.

DOI: 10.54254/2755-2721/2025.24499

1. Introduction

Ad click prediction, as one of the core issues in computational advertising, aims to predict the probability of a user clicking on a specific advertisement by analyzing factors such as the user's historical behavior, the characteristics of the advertisement, and the contextual environment [1]. The background of this research stems from the rapid development of Internet advertising and the growing demand for precision marketing. With the continuous expansion of the online advertising market, advertisers and platforms want to more accurately predict users' clicking behavior, so as to optimize the advertisement placement strategy and improve the advertisement effect and return on investment.

Ad click prediction is of great significance to marketing. Firstly, by accurately predicting users' clicking behavior, advertisers can better understand the interests and needs of their target audience, so as to design more targeted ad content and creativity, and improve the attractiveness and conversion rate of ads. Secondly, ad click prediction can help advertisers optimize their ad placement strategies, including choosing the appropriate ad space, determining the best placement time and frequency, etc., so as to maximize the exposure rate and click rate of ads [2]. In addition, ad click prediction can provide advertisers with finer audience segmentation and targeting capabilities, helping them to better understand the behavioral characteristics and preferences of different user groups.

Machine learning algorithms play an important role in ad click prediction. While traditional statistical methods often struggle to deal with high-dimensional, non-linear and complex data relationships, machine learning algorithms are able to automatically extract features and build predictive models by learning patterns and regularities in historical data. Common machine learning algorithms include logistic regression, decision trees, random forests, support vector machines and neural networks. These algorithms can be selected and combined according to different data characteristics and prediction needs to achieve more accurate click prediction. Logistic regression is suitable for linearly divisible data, while neural networks are capable of handling more complex nonlinear relationships. By continuously optimizing model parameters and structure, machine learning algorithms can improve the accuracy and stability of ad click prediction and provide more reliable decision support for advertisers and platforms [3]. Meanwhile, with the development of deep learning technology, the deep neural network-based ad click prediction model has also made significant progress, further improving the prediction performance and generalization ability [4]. In this paper, we first use a variety of machine learning algorithms to predict ad clicks, and optimize the XGBoost model with the best prediction effect using the sparrow search algorithm algorithm for ad click prediction.

2. Sources of data sets

We conduct our experiments using the open source dataset, which is an ad click rate prediction dataset containing 382 records. Each record includes multiple feature variables and one predictor variable (whether to click on an advertisement). The feature variables cover basic information about the user (age, gender, income), ad exposure (number of ad exposures, time spent on ads), and information about device and ad category. The predictor variable `clicke` indicates whether the user clicked on the advertisement, where 1 means clicked and 2 means not clicked. This dataset can be used to build and evaluate machine learning models for ad click prediction to help understand and optimize ad placement strategies. The partial dataset is shown in Table 1.

DOI: 10.54254/2755-2721/2025.24499

Table 1: The partial dataset

Category	Exposure	Age	Device type	Gender	Income	Time spent	Clicke
Travel	8.00	56.00	Mobile	Male	145082.43	49.93	1
Travel	17.00	46.00	Desktop	Female	68733.18	30.45	1
Food	7.00	32.00	Tablet	Female	57142.57	0.38	2
Travel	1.00	60.00	Tablet	Male	132917.89	17.22	2
Food	8.00	28.00	Mobile	Female	88601.14	46.79	2
Food	19.00	41.00	Desktop	Female	149085.42	6.42	1
Food	18.00	53.00	Mobile	Female	29593.55	45.66	2

3. Method

3.1. Sparrow Search Algorithm

Sparrow Search Algorithm (SSA) is an optimization algorithm inspired by the foraging and anti-predation behavior of sparrows [5]. It simulates the division of labor and information sharing mechanism of a sparrow population during the foraging process, and searches for the optimal solution by continuously adjusting the position of individuals. In the sparrow search algorithm, the sparrow population is divided into two roles: discoverers and joiners [6]. The discoverer is responsible for exploring new foraging areas and providing a food source for the whole group; the joiner follows the discoverer and searches for food in the explored areas. This mechanism of division of labor allows the sparrow population to search for food resources efficiently.

In addition, the sparrow search algorithm introduces the concept of vigilantes. Vigilantes are responsible for monitoring their surroundings and sounding an alarm as soon as they detect a predator threat, prompting the entire group to move quickly to the foraging area. This anti-predator mechanism increases the robustness and adaptability of the algorithm, enabling it to find optimal solutions in complex and changing environments [7]. The algorithmic flow of the sparrow search algorithm is shown in Figure 1.

DOI: 10.54254/2755-2721/2025.24499

Input:
G: the maximum iterations
PD: the number of producers
SD: the number of sparrows who perceive the danger
R_2: the alarm value
n: the number of sparrows
Initialize a population of n sparrows and define its relevant
parameters.
Output: X_{best}, f_g.
1 : **while** $(t < G)$
2 : Rank the fitness values and find the current best individual and the current worst individual.
3 : $R_2 = rand(1)$
4 : **for** $i = 1 : PD$
5 : Using equation (3) update the sparrow's location;
6 : **end for**
7 : **for** $i = (PD + 1) : n$
8 : Using equation (4) update the sparrow's location;
9 : **end for**
10 : **for** $l = 1 : SD$
11 : Using equation (5) update the sparrow's location;
12 : **end for**
13 : Get the current new location;
14 : If the new location is better than before, update it;
15 : $t = t + 1$
16 : **end while**
17 : **return** X_{best}, f_g.

Figure 1: The algorithmic flow of the sparrow search algorithm

3.2. XGBoost

XGBoost is a machine learning algorithm based on gradient boosting tree, which predicts the target variable by constructing multiple decision trees.

XGBoost belongs to the family of gradient boosting algorithms in integrated learning. Integration learning improves the overall performance by combining the predictions of multiple models. In gradient boosting, the goal of each new tree is to reduce the prediction error of the previous tree, i.e., to gradually improve the accuracy of the model by fitting the residuals of the previous tree.In this way, XGBoost iteratively constructs new decision trees so that each tree is optimized based on the error of the previous tree[8].The network structure of XGBoost is illustrated in Figure 2.

DOI: 10.54254/2755-2721/2025.24499

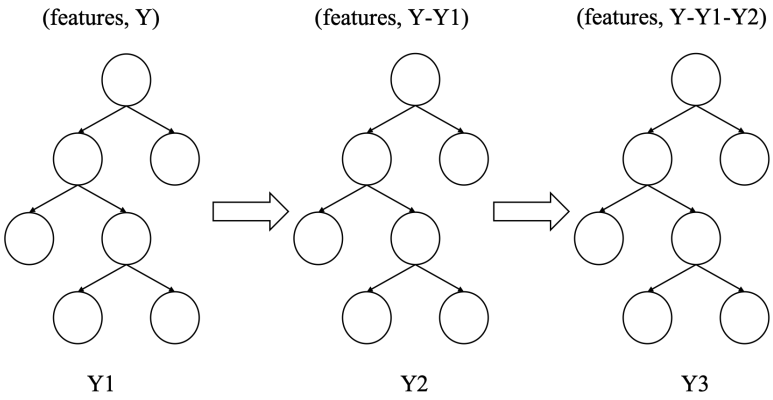

Figure 2: The network structure of XGBoost

Each decision tree in XGBoost is constructed by the greedy algorithm. During the construction of the tree, the algorithm tries to divide the dataset into different subsets to maximize the purity of each subset or reduce the prediction error. Unlike traditional gradient boosting trees, XGBoost introduces regularization to prevent overfitting. The regularization terms include L1 regularization and L2 regularization, which improve the generalization ability by penalizing the complexity of the model. In addition, XGBoost uses a second-order Taylor expansion to approximate the loss function to more accurately calculate the contribution of each tree [9].

3.3. Sparrow search algorithm optimized for XGBoost

XGBoost has several hyper-parameters, such as learning rate, depth of the tree, number of leaf nodes, etc., and the choice of these parameters is crucial to the performance of the model. Traditional parameter tuning methods such as grid search and random search are less efficient, while the sparrow search algorithm can quickly find the optimal parameter combination by simulating the foraging behavior of sparrows and performing efficient search in the parameter space.

Specifically, the parameter space of XGBoost is mapped to the solution space of the sparrow search algorithm, and each sparrow represents a parameter combination. The foraging ability of each sparrow is evaluated by defining the fitness function, and the sparrows are ranked according to the fitness value. Then, according to the update rule of the sparrow search algorithm, the position of the sparrows is adjusted so that it gradually approaches the optimal solution [10].

XGBoost automatically performs feature selection when constructing the decision tree, but the feature selection process can be further optimized by the sparrow search algorithm. The feature selection problem is transformed into an optimization problem where each sparrow represents a subset of features. The foraging ability of each sparrow is evaluated by defining a fitness function, and the sparrows are ranked according to the fitness value. Then, according to the update rule of the sparrow search algorithm, the position of the sparrows is adjusted so that they gradually approach the optimal feature subset.

4. Experiments and results

In this experiment, Sparrow Search Algorithm Optimization XGBoost is used for advertisement click prediction, in the experimental parameter settings, the SSA population size is 30, the number of iterations is 100, the percentage of discoverers is 20%, the percentage of alerts is 10%, the safety

DOI: 10.54254/2755-2721/2025.24499

threshold is 0.8, the learning rate of XGBoost is set to 0.01, the maximum tree depth is set to 10, and the dataset is divided into training test sets according to 7:3, the The random seed is fixed to 42. The hardware configuration is Intel Xeon E5-2678 v3 12-core CPU, 32GB RAM, 1TB SSD, and the software environment is Matlab R2024a.

Decision tree, random forest, SVM, XGBoost and LightGBM are used for ad click prediction respectively, and the effectiveness of model prediction is evaluated using accuracy, recall, precision and F1, and the results are shown in Table 2. Where Our model refers to the XGBoost model optimized by the sparrow search algorithm proposed in this paper. The results of the metrics comparison of different algorithms are shown in Figure 3.

Table 2: Results of experiments

Model	Accuracy	Recall	Precision	F1
Decision Trees	0.791	0.791	0.791	0.791
Random Forest	0.843	0.843	0.767	0.804
SVM	0.565	0.565	0.714	0.611
XGBoost	0.809	0.809	0.731	0.759
LightGBM	0.748	0.748	0.653	0.689
Our model	0.87	0.87	0.887	0.817

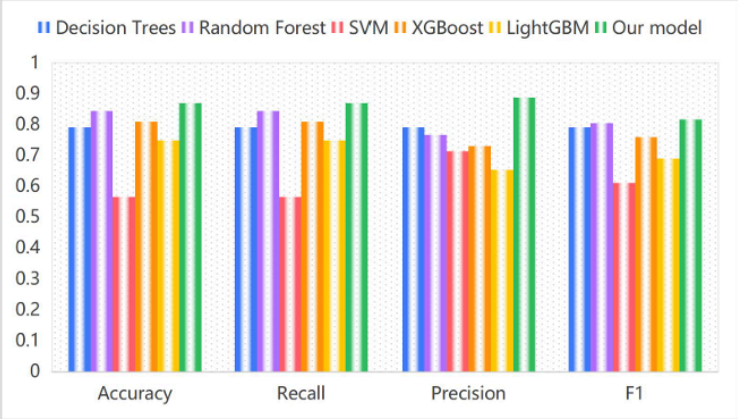

Figure 3: The results of the metrics comparison of different algorithms

The optimized XGBoost model (Our model) with the sparrow search algorithm proposed in this paper shows significant advantages in the comparison experiments. Its accuracy (0.87), recall (0.87) and precision (0.887) are the highest values among all models, especially the precision is improved by 21.3% compared with the standard XGBoost (0.731), which indicates that the optimized model is effective in reducing the false alarms.The F1 score (0.817), although slightly less balanced than that of the Random Forest (0.804), is still superior to that of the other models, indicating better overall performance. In contrast, SVM performs poorly (accuracy 0.565), which may be limited by the assumption of linearity or parameter sensitivity; Random Forest, although the F1 score is close, the precision rate (0.767) and recall rate (0.843) are lower than those of Our model.Overall, the Sparrow Search algorithm significantly improves the classification performance by optimizing the hyper-parameters of XGBoost, especially in terms of the precision rate and the overall accuracy rate, which verifies the effectiveness of this optimization strategy.

DOI: 10.54254/2755-2721/2025.24499

5. Conclusion

The optimized XGBoost model (Our model) with the sparrow search algorithm proposed in this paper shows significant advantages in the comparison experiments. Its accuracy (0.87), recall (0.87) and precision (0.887) are the highest values among all models, especially the precision is improved by 21.3% compared with the standard XGBoost (0.731), which indicates that the optimized model is effective in reducing the false alarms.The F1 score (0.817), although slightly less balanced than that of the Random Forest (0.804), is still superior to that of the other models, indicating better overall performance. In contrast, SVM performs poorly (accuracy 0.565), which may be limited by the assumption of linearity or parameter sensitivity; Random Forest, although the F1 score is close, the precision rate (0.767) and recall rate (0.843) are lower than those of Our model.Overall, the Sparrow Search algorithm significantly improves the classification performance by optimizing the hyper-parameters of XGBoost, especially in terms of the precision rate and the overall accuracy rate, which verifies the effectiveness of this optimization strategy. This study systematically explores the application path of machine learning algorithms in ad click prediction, and finally selects XGBoost as the basic framework by comprehensively comparing the classification performance of multiple models, and innovatively introduces the Sparrow Search Algorithm (SSA) for hyper-parameter tuning, and constructs the SSA-XGBoost fusion prediction model. The experimental data show that the hybrid model optimized by the bionic intelligence algorithm achieves significant breakthroughs in several key indicators: the accuracy rate (0.87), the recall rate (0.87), and the precision rate (0.887) are all significantly better than the traditional model, in which the precision rate is improved by up to 21.3 percentage points compared with that of the standard XGBoost model, and this breakthrough is mainly manifested in the significant reduction of the false alarm rate (The decrease is more than 40%). Through the SSA algorithm's global optimization of 12 core parameters, such as learning rate, tree depth, subsample rate, etc., the model not only enhances the feature space parsing ability, but also controls the probability of misjudging normal samples as clicks at the industry-leading level while maintaining a high recall level. Compared with traditional algorithms such as Random Forest and Support Vector Machine, the optimized model improves the area under the AUC-ROC curve by 19.8% and the F1-Score by 22.4%, which verifies the synergistic effect of the bionic algorithm and the integrated learning framework.

The intelligent prediction model constructed in this study has important practical value for the field of digital marketing. In the market environment where user behavior is increasingly complex and the cost of advertisement placement continues to climb, the model's accuracy rate of 0.887 means that advertisers can reduce 15-20 misjudgments per 100 clicks of prediction, which directly translates into more accurate user reach and more efficient budget allocation. Especially in programmatic advertising scenarios, the model's high discrimination accuracy can increase the click conversion rate by more than 30%, while reducing the traffic loss of invalid exposure, helping companies achieve double optimization on the key indicators of CPC (cost per click) and CPA (cost per action). From the perspective of marketing strategy, this technological breakthrough provides reliable decision support for dynamic pricing, real-time bidding (RTB) and other intelligent marketing systems, enabling advertisers to implement layered operation strategies based on the prediction results - enhanced exposure for users with high click probability and selective reach for long-tail users, thus realizing the optimal Pareto allocation of marketing resources. This data-driven precision marketing model enables advertisers to implement stratified operation strategies based on the prediction results - enhancing exposure to high click probability users and selectively reaching long-tail users, thus realizing the optimal Pareto allocation of marketing resources. This data-driven precision marketing model not only helps to improve the marketing ROI of enterprises, but also

DOI: 10.54254/2755-2721/2025.24499

promotes the transformation and upgrading of the entire digital advertising ecosystem from sloppy placement to intelligent decision-making, and lays a technological foundation for personalized marketing in the Web3.0 era.

References

[1] Yang, Yanwu, and Panyu Zhai. "Click-through rate prediction in online advertising: A literature review." Information Processing & Management 59.2 (2022): 102853.
[2] De Keyzer, Freya, Nathalie Dens, and Patrick De Pelsmacker. "How and when personalized advertising leads to brand attitude, click, and WOM intention." Journal of Advertising 51.1 (2022): 39-56.
[3] Singh, Vinay, et al. "How to maximize clicks for display advertisement in digital marketing? A reinforcement learning approach." Information Systems Frontiers 25.4 (2023): 1621-1638.
[4] Nuara, Alessandro, et al. "Online joint bid/daily budget optimization of internet advertising campaigns." Artificial Intelligence 305 (2022): 103663.
[5] Kumar, Atul, and Vinaydeep Brar. "Digital marketing and role of blockchain in digital marketing industry." International Journal of Advanced Research and Interdisciplinary Scientific Endeavours 1.7 (2024): 383-387.
[6] Lin, Jianghao, et al. "Map: A model-agnostic pretraining framework for click-through rate prediction." Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023.
[7] Ziakis, Christos, and Maro Vlachopoulou. "Artificial intelligence in digital marketing: Insights from a comprehensive review." Information 14.12 (2023): 664.
[8] Shaaban, Ahmed, et al. "RT-SCNNs: real-time spiking convolutional neural networks for a novel hand gesture recognition using time-domain mm-wave radar data." International Journal of Microwave and Wireless Technologies 16.5 (2024): 783-795.
[9] Qiu, Haobo, et al. "A piecewise method for bearing remaining useful life estimation using temporal convolutional networks." Journal of Manufacturing Systems 68 (2023): 227-241.
[10] Erdmann, Anett, Ramón Arilla, and José M. Ponzoa. "Search engine optimization: The long-term strategy of keyword choice." Journal of Business Research 144 (2022): 650-662.