

# ***Research on Advertising Prediction Based on Stacking Ensemble Algorithms***

**Sirui Fang<sup>1,a,\*</sup>**

<sup>1</sup>*School of Computer Science and Engineering, North Minzu University, Ningxia, China*

*a.fsr144603@outlook.com*

*\*corresponding author*

**Abstract:** The purpose of advertising prediction is to analyze user behavior and features to forecast the effectiveness of ad placements, thereby achieving a higher return on investment. This paper employs a stacking-based ensemble learning method, integrating multiple machine learning algorithms to enhance predictive capabilities. To support this method, this article selected a dataset related to advertising click through rate (CTR). The paper constructs two models: Gradient Boosting Decision Tree (GBDT) and eXtreme Gradient Boosting (XGBoost). Then the paper utilizes Bayesian optimization to tune the model parameters to achieve optimal parameter combinations. The paper uses these two individual models as base learners in a stacking ensemble model, with a random forest model serving as the meta-learner for training. This paper evaluates the predictive performance of the models using two metrics: log loss value, and AUC value. After experimental evaluation, the AUC value of the Stack fusion model reached 0.9999, which is about 0.42%, 1.29%, and 10.99% higher than XGBoost (0.9956), Random Forest (0.9871), and GBDT (0.8904), respectively. In terms of logloss value, the Stack fusion model has a loss of 0.0101, significantly lower than XGBoost's 0.0595, a decrease of about 83.01%. Compared to the random forest's 0.1896 and GBDT's 0.1884, it decreased by approximately 94.67% and 94.64%, respectively. This result indicates that the ensemble model proposed in this paper demonstrates a significant improvement in click-through rate prediction when contrasted with individual models.

**Keywords:** Click through rate, GBDT, XGBoost.

## **1. Introduction**

The global advertising market has shown a continuous expansion trend, becoming a key force driving the development and competition of various industries. With the continuous advancement of digital technology, the scale of the advertising market is also steadily growing. This is particularly evident in the field of online advertising, where the widespread adoption of mobile internet and smart devices has diversified and enriched the forms and channels of advertising. These factors have promoted the rapid growth of the global advertising market, making it an indispensable part of the global economy.

Media investment company Magna predicts global media advertising revenue will reach \$927 billion in 2024. With the rise of social media platforms (such as Facebook, Instagram, and TikTok) and search engine advertising (such as Google), advertisers can achieve more efficient marketing goals through accurate user data analysis and delivery strategies.

The broad prospects of the advertising market have attracted a lot of research and technological innovation, especially in terms of accuracy and effect evaluation of advertising. Traditional advertising prediction models usually depend on simple rules or historical data analysis, and it is difficult to cope with the huge and complex characteristics of modern advertising data. In this context, artificial intelligence have gradually become the core driving force for the advertising industry.

Especially the research based on stacking fusion algorithms is becoming an emerging advertising prediction method. The stacking algorithm can be weighed and fused by the prediction results of multiple basic models, and further improve the predictive accuracy of the indicators such as advertising clicks. To cope with the massive characteristics of the advertising data, Xiao Xiaoling and others proposed an innovative Light Gradient Boosting Machine (XGB\_LGBM) model that combines the combination of mutual information features and stacking algorithms, which greatly improves the predictive effect of advertising clicks [1]. Lu proposed a Stacking ensemble classification algorithm AVWNB Stacking based on attribute value weighted naive Bayes algorithm, utilizing the prior information obtained from the Stacking learning mechanism and the powerful probability expression ability of Bayesian networks. This algorithm effectively improves the classification performance of the model [2].

In recent years, many scholars have proposed a variety of machine learning models to improve advertising effects. For example, Zhang feeding neural networks (FNN) models, use deep learning methods to effectively improve the predictive ability of advertising clicks, showing the huge potential of deep learning in the advertising field [3]. On the basis of the GBDT model, Wang has significantly improved the model's prediction accuracy and computing speed by replacing it with XGBOOST and combined with the FFM model [4]. Zheng put forward a GBDT-Deep AFM model through a fusion gradient to improve the decision tree (GBDT), deep factor decomposition machine (Deep FM), and attention mechanism, and tapped the deep-level information in the advertising data more deeply [5]. Similarly, Deng combined the Gaussian hybrid model (GMM) with factor decomposer (FM) to build the GMM-FMS advertising click-through rate prediction model, which achieved a more ideal predictive effect [6].

Li's research suggests that introducing the fusion of cross networks and deep neural networks can enhance the performance of the stacking fusion model. This approach has shown potential for improving prediction accuracy and model stability in similar scenarios, effectively addressing issues of data sparsity and information overload. In the process of exploring ways to improve the performance of stacked fusion models, Li proposed a valuable viewpoint that introducing the fusion of cross networks and deep neural networks can significantly improve the model's performance. This improvement has demonstrated the potential to improve prediction accuracy and model stability in similar scenarios, while effectively addressing the issues of data sparsity and information overload [7].

These research results demonstrate the complexity of advertising data and its multi-vita signs, reflect the broad needs of the advertising market, and provide important technical support for the optimization of future advertising strategies. With the continuous development of technology, the advertising field is moving towards a more efficient and smarter direction. Machine learning and artificial intelligence will continue to play a vital role in advertising forecasting, accurate release, and effect assessment.

In the background of the current research, this paper employs the Stacking algorithm. By integrating Bayesian optimization, the model is refined to select optimal hyperparameters, thereby enhancing performance. The study utilizes XGBoost and GBDT as base learners, with Random Forest serving as a component learner. Compared with existing research, this article has innovated in the strategy of optimization and model fusion of models, which aims to achieve better prediction effects.

## 2. Related algorithms

### 2.1. Gradient boosting decision tree (GBDT)

GBDT is an iterative ensemble learning algorithm widely used for classification and regression problems. The fundamental idea of GBDT is to construct multiple decision trees incrementally, where each tree is optimized based on the prediction errors of the previous tree. In each iteration, the model trains a new decision tree by fitting the residuals (the differences between the actual and predicted values). Ultimately, the predictions from multiple trees are weighted and summed to produce the final output. The advantage of GBDT lies in its ability to handle complex data patterns, often resulting in excellent predictive accuracy. The prediction formula for the model is as follows:

$$F(x) = \sum_{m=1}^M h_m \quad (1)$$

### 2.2. Extreme gradient boosting (XGBoost)

XGBoost is an optimized version of Gradient Boosting Decision Trees (GBDT) that focuses on enhancing model training efficiency and predictive accuracy. Proposed by Chen in 2016, XGBoost has rapidly gained widespread application in the field of machine learning [8]. Its core principle involves incrementally constructing multiple decision trees and aggregating their outputs into a strong learner. Compared to traditional GBDT, XGBoost incorporates various optimization techniques, including parallel computation, regularization, and the handling of missing values, which enable it to perform exceptionally well on large-scale datasets. Furthermore, XGBoost allows for the evaluation of feature importance, aiding in feature selection for the model. Building upon GBDT, XGBoost introduces additional regularization terms, and its objective function is defined as follows:

$$Obj(\theta) = \sum_{i=1}^n l(y_i, F(x_i)) + \sum_{j=1}^K \Omega(f_j) \quad (2)$$

$$\Omega(f_j) = \gamma T + \frac{1}{2} \lambda ||w_j||^2$$

The final prediction of XGBoost is also represented as a summation of the trees:

$$F(x) = \sum_{i=1}^K f_i(x) \quad (3)$$

### 2.3. Random forest (RF)

Random Forest is an ensemble learning algorithm that constructs multiple decision trees using the "Bagging" approach. During the training of each tree, subsets are randomly sampled from the original data, and a random selection of features is chosen for splitting at each node. This method effectively reduces overfitting and enhances the model's generalization capability. Random Forest is particularly suitable for high-dimensional feature data and does not require dimensionality reduction, resulting in excellent performance in various practical applications. The prediction of the Random Forest model can be expressed as:

$$F(x) = \operatorname{argmax} \sum_{i=1}^K f_i(x) \quad (4)$$

### 2.4. Bayesian optimization

Bayesian optimization is an efficient method for optimizing black-box functions, particularly suitable for hyperparameter tuning. In machine learning, selecting appropriate hyperparameters is crucial for model performance; however, the search space for hyperparameters is often complex and computationally expensive. Traditional methods such as grid search or random search may not effectively find the optimal solution. Bayesian optimization constructs a probabilistic model of the

objective function and intelligently selects the next test point, allowing it to find near-optimal hyperparameter configurations in fewer iterations, thereby saving computational time and exhibiting strong global search capabilities. Although the effectiveness of Bayesian optimization depends on the assumptions of the chosen model and may face challenges in high-dimensional scenarios, its potential applications in tuning complex models and large-scale datasets are extensive. When applied judiciously, it can significantly enhance model performance and training efficiency.

### 3. Experimental process

This experiment aims to explore the optimization of the advertising prediction model, and use the experimental steps of a series of systems to verify the effectiveness of the Stacking fusion model in advertising prediction. As shown in Figure 1, the flow chart of the Stacking model shows the logical relationship between each step:

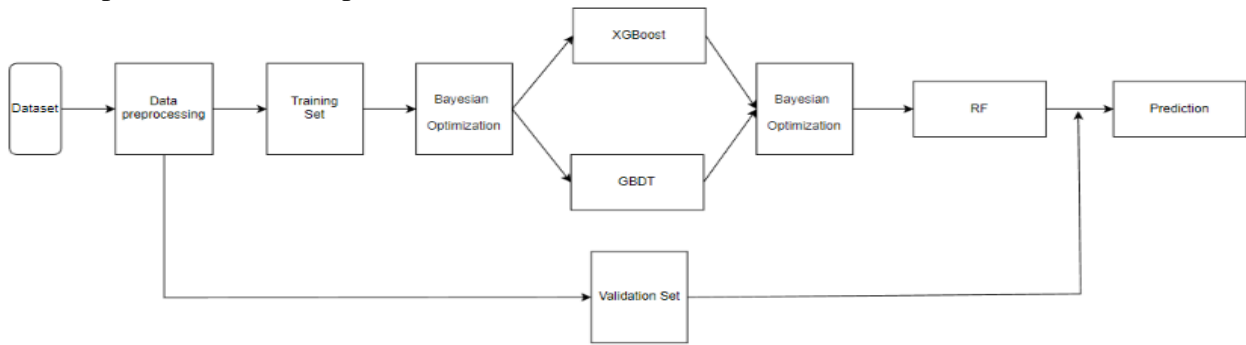


Figure 1: The structure of the autoencoder (Photo/Picture credit: Original).

#### 3.1. Experimental dataset

The dataset used in this study is sourced from Kaggle and focuses on the click-through rate (CTR) of advertisements, containing 100,000 samples. This public dataset, contributed by global data scientists and researchers, has been widely used and validated, ensuring high reliability. It is designed to analyze and model the relationship between ad impressions and user clicks behavior, enabling better predictions of future advertisement click-through rates. In this study, 70% of the dataset is allocated for the training set and 30% for the validation set to evaluate the model's performance.

#### 3.2. Model parameter optimization

This paper determines the related parameters of the model through Bayeste optimization to improve the performance and predictive ability of the model. Tables 1, 2, and 3 respectively provide the related parameters of GBDT, XGBoost, and RF after Bayesian optimization:

Table 1: GBDT parameter configuration

Model	Parameter name	Parameter value
GBDT	n_estimators	60
	max_depth	10
	learning_rate	0.1161

Table 2: XGBoost parameter configuration

Model	Parameter name	Parameter value
XGBoost	n_estimators	200
	max_depth	5
	learning_rate	0.3
	colsample_bytree	0.5623
	subsample	0.8485

Table 3: RF parameter configuration

Model	Parameter name	Parameter value
RF	n_estimators	57
	max_depth	9
	min_samples_split	2
	min_samples_leaf	8

#### 4. Experimental results

The AUC value is an important metric for evaluating a model's distinguishing ability, reflecting the probability that a randomly chosen positive sample ranks higher than a randomly chosen negative sample. Figure 2 shows the comparison of AUC values among GBDT, XGBoost, Random Forest (RF), and Stacked Fusion models. AUC (area under the curve) is an indicator to measure the performance of a classification model, and the closer the value is to 1, the stronger the classification ability of the model. The result shows that the Stacking ensemble model exhibits the highest AUC value, indicating its superior predictive performance. In contrast, the AUC value for GBDT is relatively low, suggesting that its predictive performance is inferior compared to the other models.

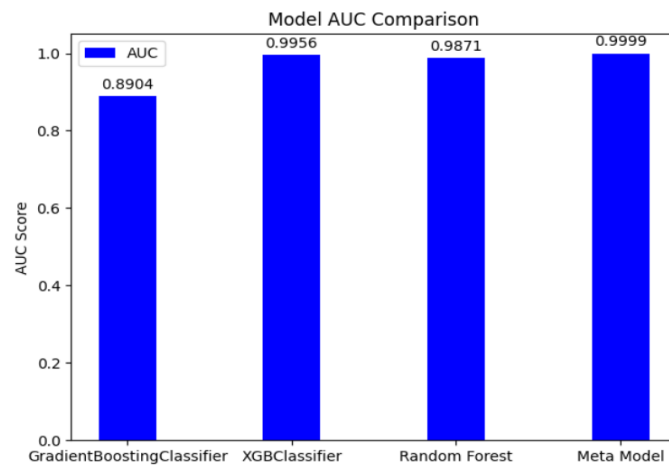


Figure 2: Comparison results of AUC values for different models (Photo/Picture credit: Original).

Log Loss is a commonly used loss function, which is suitable for the classification problem of prediction probability. It helps evaluate the performance of the model by quantifying the difference between the predicted probability and the actual label. As shown in Figure 3, the Log Loss value of the Stacking ensemble model is significantly lower than that of the other models, indicating that it is better in predictive accuracy.

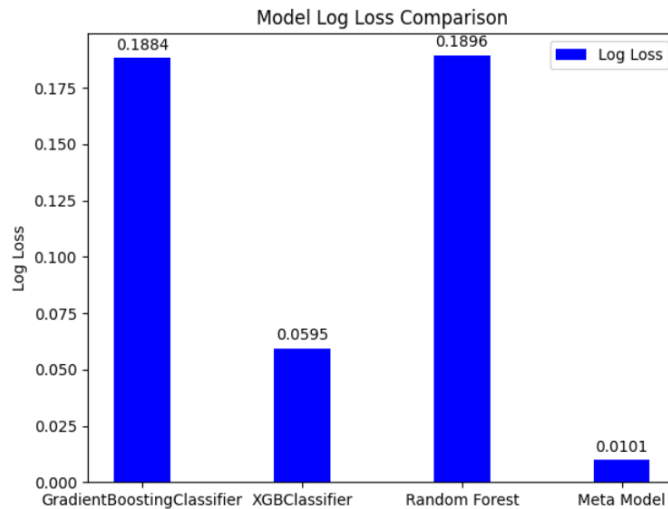


Figure 3: Comparison results of log loss values for different models (Photo/Picture credit: Original).

## 5. Conclusion

This paper employs Bayesian optimization to determine the optimal parameters for extracting product features. Building on this foundation, a Stacking ensemble model based on GBDT, XGBoost, and RF is established using the processed dataset for ad prediction. The proposed model is compared with individual models. The results indicated that the stacking fusion model performed significantly better than other models in AUC value and logloss indicators, indicating its strong adaptability in advertising prediction, which is consistent with the results shown in the figure.

Although this study has made some progress in model design and performance evaluation, there are still many shortcomings. Firstly, the current experiment did not use cross validation methods to evaluate the model's generalization ability, which may lead to overly optimistic estimates of the model's performance. Cross validation can effectively avoid overfitting of the model to the training set, thereby improving its predictive ability on unknown data. Therefore, future research should consider introducing techniques such as k-fold cross validation to obtain more reliable performance evaluations. In addition, the selection and combination of base learners is also a key issue in the application of ensemble learning. Subsequent work can explore more diverse base learner combination strategies to further improve the accuracy and robustness of the model. Improvements to address these shortcomings will provide a solid foundation for future research.

The main limitation of this study is the limited diversity and scale of the dataset, and the current dataset used may not fully represent all scenarios of advertising placement, which may result in some results not fully reflecting the actual situation.

Future research direction: In response to the above shortcomings, future research can consider introducing more diverse datasets, exploring new deep learning models, or conducting more in-depth hyperparameter tuning to improve the model's generalization ability and practical application value.

## References

- [1] Xiao, X., Li, X., & Liu, T. (2022). *Research on improved CTR prediction model*. *Journal of Yangtze University (Natural Science Edition)*, 1-7.
- [2] Lu, W., Xu, J., & Li, Y. (2022). *Improved classification algorithm for stacking integration and its applications*. *Computer Applications and Software*, 39(2), 281-286.
- [3] Zhang, W., Du, T., & Wang, J. (2016). *Deep learning over multi-field categorical data*. In *European Conference on Information Retrieval* (pp. 45-57). Springer, Cham.
- [4] Wang, X. (2018). *CTR estimation based on RF, XGBoost, and FFM integration (Master's thesis)*. Zhejiang University.
- [5] Zheng, Z. (2023). *Research and application of advertising clicks based on GBDT and attention mechanism (Master's thesis)*. Nanchang University.
- [6] Deng, L., & Liu, P. (2019). *GMM-FMS-based advertising click-through rate prediction research*. *Computer Engineering*, 45(05), 122-126.
- [7] Li, B. (2024). *Research and application of personalized recommendation model based on deep learning (Master's thesis)*. Linyi University.
- [8] Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. *ACM*.