Digital Management Mode of Real Estate Marketing based on Big Data and Artificial Intelligence

SHUANGXIN CHEN
Lyceum of the Philippines University Manila Campus,
Manila 1002,
PHILIPPINES

Abstract: - To cope with the pressure on sales information processing as the real estate industry grows, the study builds a real estate digital marketing management system design based on the analysis of real estate marketing needs to meet the needs of real estate marketers for digital information processing, and builds a hybrid recommendation model using a combination of Gradient Boosting Decision Tree (GBDT) technology and Logistic Regression (LR) to accurately recommend real estate potential purchase users. The GBDT-LR model performance test results show an accuracy of 94.63% and a regression rate of 94.82%, which is particularly good in terms of classification accuracy, and the system CPU occupancy rate basically stays below 30% during the whole script running period, and the system still maintains good system stability when the TPS user concurrency is 150, and it's using experience is better. The comparison of the ROC curve of the GBDT-LR model shows that the GBDT-LR model's accuracy is as high as 92%, which is better than the performance of most of the classification models, and it can meet the practical application requirements of the real estate industry and provide a good solution for the real estate industry. It can meet the actual application requirements of the real estate industry and provide a scientific and systematic digital management solution for the real estate industry.

Key-Words: - Gradient Boosting Decision Tree, Logistic Regression, Marketing management, Machine algorithm, real estates, digital management.

Received: March 26, 2023. Revised: November 12, 2023. Accepted: February 14, 2024. Published: March 29, 2024.

1 Introduction

As one of the important pillars of the global economy, the marketing management model of the real estate industry has been the focus of attention in the industry and academia. The traditional real estate marketing model relies heavily on human resources and experience, but with the rapid development of technology, especially the rise of big data and artificial intelligence algorithms, this model is facing tremendous pressure for change. Big data technology can extract useful information from massive amounts of data, providing real estate marketing with more accurate target customer positioning, price optimization and market trend forecasting. Meanwhile, artificial intelligence technologies, especially machine learning and natural language processing, are gradually changing the face of real estate marketing, [1], [2]. For example, AI can automatically adjust marketing strategies by analyzing customer behavior and feedback to improve conversion rates. However, despite the huge potential of big data and AI technologies, how to effectively integrate these advanced technologies into real estate marketing and how to build an efficient and sustainable digital management model remains an unresolved issue. Currently, most real estate companies' attempts in this regard are mostly sporadic and localized, lacking a comprehensive and systematic application framework. In addition, the promotion and application of digital management models face multiple challenges, including data security, user privacy protection, and compatibility with existing systems, [3]. Therefore, how to effectively implement digital management while ensuring the interests and compliance of all parties has become an urgent issue. The study aims to delve into the digital management model of real estate marketing based on big data and artificial intelligence. By analyzing existing research and practices, as well as conducting in-depth studies on multiple case studies, the study attempts to build a comprehensive, efficient, and sustainable digital management model. The study expects to provide a scientific and systematic digital management solution for the real estate industry, as well as a

reference for relevant policy-making and future research.

2 Related Works

Research on real estate has focused on a review of the residential characteristics of real estate. Focusing on how the form/characteristics of compact/hilly public housing communities affect dementia in Asian older adults, [4], conducted a cross-sectional analysis of 2,077 elders living in public housing estates in Hong Kong, measuring dementia through the Montreal Cantonese version of the Cognitive Assessment. The built environment was measured according to three dimensions (greenery, walkability, and accessibility) and included 11 indicators. Results suggest that not considering the form/characteristics of walking paths may overestimate the health benefits of the built environment. Focusing on how the COVID-19 pandemic has changed the motivations and housing preferences of investors in the property market, [5], explored residential preferences in terms of quantity, spatial extent, and relationship to social infrastructure, statistically analyzing the unit prices of sold properties. The case studies show a marked increase in demand for residential properties away from the parts of the city with the highest density of social infrastructure, favoring areas on the urban fringe and close to green spaces. Artificial intelligence algorithms have applications in various fields, [6], to solve the problems of low accuracy and slow speed of traditional coal gangue recognition methods, proposed an intelligent classification method of coal gangue using YOLOv5 and multispectral imaging technology. Experimental results show that the average accuracy of gangue detection using the YOLOv5.1 model reaches 98.34%, which can not only accurately identify the gangue, but also obtain the relative position of the gangue, which can be effectively used for the identification of coal gangue. Han et al. combined recurrent neural networks and LSTM network to construct a system to predict dynamic gestures by joint coordinate features. In the experiments, the model achieves the highest accuracy of 99.31%, indicating superior recognition performance, [7]. [8], proposed a YOLOv5-based motorbike helmet detection method for motorcyclist helmet detection via video surveillance, which uses soft-NMS instead of NMS to fuse the YOLOv5 detector, and experimentally achieves 97.7% of mAP, 92.7% of F1 scores, and 63 frames-per-second (FPS), which is better than other state-of-the-art detection methods. [9], focused on automated planning and cost estimation of concrete formwork, and to accomplish the automatic generation of bills of materials (BoMs) for formwork, a BoM generation AI model based on Mask R-CNN and image segmentation techniques (BoM-GAIM) was proposed. The model can identify, classify, and extract formwork components with an accuracy of up to 98%, and when integrated with the cost database, BoM-GAIM can generate BoMs for concrete formwork in a user interface environment, which improves design efficiency. [10], proposed an sEMG gesture recognition model consisting of feature extraction, genetic algorithm (GA), and support vector machine (SVM) model for accurately distinguishing different surface electromyography (sEMG) gestures for intelligent prosthetic limb control, and used the adaptive mutation particle swarm optimization (AMPSO) algorithm to optimize the parameters of SVM. The results show that the sEMG gesture recognition rate is 0.975 for AMPSO-SVM, 0.9463 for PSO-SVM, 0.9093 for GS-SVM, and 0.9019 for BP, which can effectively recognize the low-frequency sEMG signals with different gestures. Ezaldeen et al. used the NPSO algorithm to learn the importance of the types of relationships between concepts to complete a simulated recommender system based on the highest rankings for dynamic learners for both the CLM and ECLM conceptual models and the results of the simulation proved that the ECLM performs better than the other existing methods, with a mean reciprocal rank (MRR) value of 0.780, [11]. [12], proposed a hybrid technique for energy management systems (EMS) between electric vehicles (EV) and distribution systems. The proposed hybrid system jointly performs the Fertiliser Field Algorithm (FFA) and Gradient Boosted Decision Tree (GBDT). The performance test results show that the proposed technique is effective in finding the near-global optimal solution with less computation, while the energy consumption of the technique is 720.34 KJ, which is lower than the existing algorithms. Unlike the existing literature, the study not only focuses on real estate markets and residential properties but also provides insights into how to improve the accuracy and reliability of the system through advanced hybrid GBDT and logistic regression (LR) models. Through comprehensive performance testing and data analysis, the study provides strong technical support for real estate digital transformation.

3 Design of Real Estate Digital Marketing Management System Based on GBDT Algorithm

The research for the design of digital management aspects applicable to real estate marketing, on the one hand, based on the analysis of real estate marketing needs to build a real estate digital marketing management system design to meet the needs of real estate marketing personnel digital information processing; on the other hand, the use of Gradient Boosting Decision Tree (GBDT) technology and Logistic Regression (LR) technology to build a hybrid recommender model, to accurately recommend the potential purchase of real estate users.

3.1 Real Estate Digital Marketing Management System Design and Module Function Division

As urbanization accelerates in China, demand for real estate is growing exponentially. This demand is based on multi-level and multi-dimensional considerations, with the most basic level being the demand for the basic residential function of housing. However, in the process of urbanization, the demand for real estate is also diversified and stratified due to the increased mobility of the population and the complexity of social classes. The housing demand of the general public mainly focuses on the purchase of houses for marriage, house demolition and relocation, employment and settlement, as well as home purchase, etc. In addition to the requirements for the basic attributes of houses, such as area and house type, the accessibility to transport, commercial facilities, medical resources and the quality of school districts have also become the key factors influencing the demand for property. These comprehensive demands not only influence consumer choices but also provide the basic framework for a property marketing management system. Therefore, an efficient and complete property marketing system needs to be able to comprehensively reflect and satisfy these diversified needs. The functional division of the real estate digital marketing management system designed by the study is shown in Figure 1.

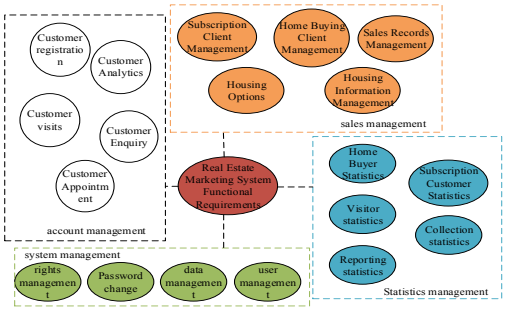

Fig. 1: Functional division of real estate digital marketing management system

Real estate digital marketing management system covers all aspects of real estate sales, mainly divided into four parts: customer management, sales management, statistics management, and system management. In the real estate marketing digital management system, the customer relationship management module occupies a core position. This module focuses on the comprehensive integration and maintenance of customer information, including but not limited to customer-level classification, appointment information, and visit records. By building a highly structured database, the system achieves the integration of multi-dimensional information such as name, phone number, the content of inquiry and intention to purchase a home. In particular, the information on home purchase intention and intended properties can be dynamically updated to adapt to the rapid changes in the market and customer needs. The design of this module needs to fully consider the four paradigms of databases to ensure data consistency and maintainability. Another key module is Property Management, which aims to achieve integrated management of property information, including detailed records on several aspects such as project house type, floor, and location. By comprehensively analyzing these data, the system can update and provide real-time property information that best meets customers' needs. The sales management module pays more attention to the refined management of the sales process. It includes several sub-modules such as coordinator information registration, successful purchase customer information management, and sales performance information management. This not only improves the accuracy of sales but also tracks the performance of sales personnel in real-time, providing strong support for the company's decision-making. Statistical management, i.e., the decision support module, through big data analysis, conducts comprehensive analysis of multi-dimensional information such as the strength of customers' purchase intention, market trends, and sales

performance, thus assisting the company's leadership to make smarter decisions in the complex market environment. Finally, the system background management module is mainly operated by the system administrator, who is responsible for the stable operation and maintenance of the whole system. The module includes key functions such as user information management, role management, permission setting, and data backup and recovery to ensure that the system can continue to serve the real estate sales business stably. Based on the above analysis, there is the overall functional design of the system as shown in Figure 2.

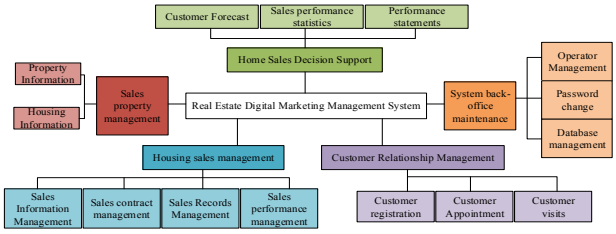

Fig. 2: Overall functional design of the system

Real estate digital marketing management systems can collaborate to deal with the overall sales of real estate, which is divided into five modules, respectively, for the sale of property management, housing sales management, customer relationship management, housing sales decision-making support and system background maintenance. The property management module is mainly responsible for the overall management of housing sales information, including home buyers are concerned about the building, unit number, number of floors direction of the house type, and other information, as well as the project's location, housing, commercial and living facilities and so on. Information processing and management are mainly achieved through operations such as adding, deleting, modifying, and querying. In the design process, object-oriented programming methods are adopted and Property Model is introduced as the encapsulated object of information management. The salesman registers all kinds of information in the system, such as orientation, house type, unit price, etc., and is subject to the system's compliance checking The decision support submodule, on the other hand, supports corporate decision-making by statistically analyzing the customer's demand for purchasing properties. The background management of the system is divided into three sub-modules: user management, role management, and permission management. User management mainly involves the addition, modification, deletion, and query of account information; role management includes the addition, modification, and deletion of roles and the setting of user roles; and permission management is responsible for setting the corresponding permissions of roles. Customer relationship management, as the core module of the system, needs to complete the collation of real estate enterprises of various types of customer demand information and relationships, and its customer information query process is shown in Figure 3.

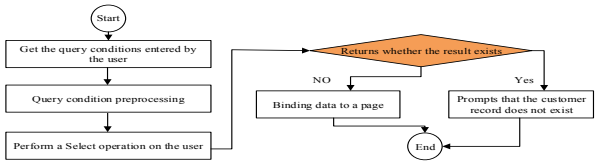

Fig. 3: Customer information inquiry process

As shown in Figure 3, after the customer information management module obtains any customer information such as age, number of times of purchase, name, etc. entered by the user, the database preprocesses the query conditions and presents a table of customer service information that meets the selection of the target customer information table to return the results, the customer exists then binds the data to the page, and if the customer does not exist then the customer is prompted to indicate that no record of this customer exists. In addition to this software design should also consider the system performance, usability, and scalability of the system's non-functional requirements, first of all, the interface and multi-terminal adaptation requirements are considered an important factor in the user experience. The system utilizes the latest standards of HTML5 and CSS3 and adopts a responsive design approach to ensure consistency and usability across the diversity of end devices. Secondly, as the system needs to handle a large amount of sales and customer relationship data, special emphasis was placed on computing performance in the design. The back-end architecture adopts a distributed computing mechanism, and load balancing and data slicing technologies to meet the needs of large-scale data processing and high-speed computing. Third, system stability and robustness is another key consideration. By applying microservice architecture and containerization technology, the system achieves a high degree of availability and failover capability. Finally, the system also gives full consideration to scalability in its interface design, adopting RESTful API and OpenAPI standards to support possible third-party integration or functional upgrades in the future. These design strategies not only meet the current needs of the system but also provide a basis for its long-term

maintainability and scalability. The relationship between the database entities is shown in Figure 4.

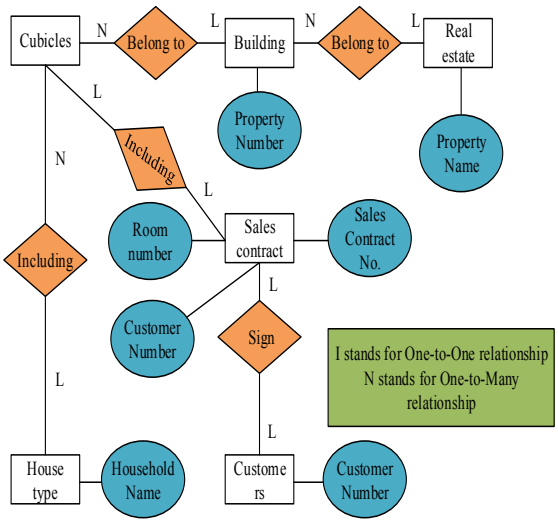

Fig. 4: Relationships between entities in the database

In database design, the Third Normal Form (TNF) is often considered an idealized design standard aimed at eliminating data redundancy. However, in practice, over-reliance on the Third Paradigm may lead to complex table structures and reduce the efficiency of data manipulation. Therefore, the system proposed in the study is designed to appropriately relax the paradigm requirements to improve the performance of database querying and updating by introducing limited redundant data. The system adopts a Model First-based design approach, where database tables and fields and the associations between them are automatically generated mainly through entity mapping. This design pattern emphasizes the design of entity models and entity attributes. Firstly, based on the requirements analysis and the data flow needs of the functional modules, the study was conducted to classify and integrate the entities according to a standardized process. This yields all the core entity models required by the system, including users, house types, floor plans, rooms, and customers. Secondly, in the selection of attribute types, in addition to considering resource consumption and performance optimization (e.g. try to use int type and avoid varchar type), the variability of attribute value types in practical application scenarios also needs to be taken into account.

3.2 Design of Real Estate Recommendation Algorithm based on GBDT and LR

Currently, the identification and development of high-value customers has become the core policy of real estate marketing, which is oriented to customer needs to enhance market influence. However, the traditional sales approach lacks in-depth analysis of consumer behavior and data mining, which can easily lead to inconsistency between the company's promotional strategy and customers' needs, affecting precise marketing. Therefore, it becomes imperative to establish customer value, find potential consumers, and provide theoretical support for personalized marketing strategies. The study will construct a potential customer identification model based on user-generated content to identify property potential customers to make accurate marketing recommendations, the study chooses to use a hybrid recommendation algorithm based on the fusion of Gradient Boosting Decision Tree (GBDT) and LR (Logistic Regression). This algorithm integrates collaborative filtering, housing system filtering, and matrix-based collaborative filtering. The main innovation of this fusion model is that it introduces the idea of an advertisement recommendation algorithm and takes the user's click information as the input of the model to improve the recommendation accuracy. The GBDT algorithm has low computational complexity and excellent fault tolerance, which makes it especially suitable for dealing with nonlinear and noisy data. Compared with a single decision tree, GBDT integrates multiple decision trees through gradient boosting, which effectively mitigates the overfitting problem, and thus provides more accurate and reliable recommendation outputs for potential customer identification, [13], [14], [15]. The LR model is based on the linear regression model and applies the Sigmoid function, which makes the results of the linear regression model fall on [0, 1], and the linear regression function is shown in Eq. (1).

$$h(x) = \theta^T x = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n \quad (1)$$

In Eq. (1), both denote matrices; θ is the linear regression parameter, and denotes the input features. The sigmoid function is shown in Eq. (2).

$$y(x) = \frac{1}{1+e^x} \quad (2)$$

The Sigmoid function is a smooth curve, centrosymmetric about (0, 0.5), and the Sigmoid function is usually used for mapping the value domain, and the LR model is constructed based on linear regression and the Sigmoid function, which is shown in Eq. (3).

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}} \tag{3}$$

The LR model has a wide range of applications in distributed environments with its powerful parallel processing capability and simple model structure, but its learning capability is limited by the need to obtain a large number of valid feature combinations in advance to enhance its nonlinear learning capability. GBDT is an integrated algorithm based on a decision tree, and the core idea is to reduce the residuals of the prediction results from the previous tree by continually fitting the residuals of the prediction results from the previous tree. The approximation of the residuals in each round is fitted by the negative gradient of the loss function, i.e., the CART regression tree is generated, and the negative gradient is represented in Eq. (4)

$$r_{ki} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)-f_{k-1}(x)} \tag{4}$$

r_{ki} denotes the negative gradient of the loss function of the i th sample of the k th round; L denotes the loss function; ∂ denotes the fitting parameter $f(x_i)$ denotes the loss function of the i th sample; the use of $(x_i, r_{ki}), i = 1, 2, \cdots, m,$ calculations can be fitted to a CART regression tree, for the k th regression tree; there are corresponding leaf node regions $R_{kj}, j = 1, 2, \cdots, J$; J denotes the number of leaf nodes, each leaf node samples have a loss function of the minimum output value calculation see Eq. (5).

$$C_{kj} = \arg\min \sum_{x_i \in R_{kj}} L(y_i, f_{k-1}(x_i) + C) \tag{5}$$

In Eq. (5), C_{kj} denotes the minimum output value and C denotes the residual fitting parameter, the corresponding decision tree fitting function for this round is shown in Eq. (6)

$$h_k(x) = \eta \sum_{j=1}^{J} C_{kj} I(x \in R_{kj}) \tag{6}$$

In turn, there is a strong learner expression see Eq. (7)

$$f_k(x) = f_{k-1}(x) + \eta \sum_{j=1}^{J} C_{kj} I(x \in R_{kj}) \tag{7}$$

GBDT is often used to solve classification problems and regression problems, hybrid recommendation algorithms are solved using the GBDT classification algorithm. GBDT classification algorithm long using log-likelihood loss function, the algorithmic process is, first of all, the model initialization, the computational formula is shown in Eq. (8).

$$f_0(x) = \arg\min \sum_{i=1}^{N} L(y_i, f_0) = \log \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} (1-y_i)} \tag{8}$$

The loss function is shown in Eq. (9)

$$L(y_i, f_0(x_i)) = -\sum_{i=1}^{N} (y_i \log p_i + (1-y_i)\log(1-p_i)), p = \frac{1}{1+e^{-f_0(x_i)}} \tag{9}$$

The second step calculates the negative gradient of the loss function for the i th sample of the k th round, calculated as in Eq. (10).

$$r_{ki} = -\eta \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)-f_{k-1}(x)} \tag{10}$$

In the third step, the optimal cut-off variable and the optimal cut-off point are selected, and the minimum output value of the loss function is calculated, as shown in Eq. (11).

$$\min_j \left[\min_m \sum_{x_i \in R_{1j}^m} (r_{k,i} - C_{1j}^{(k)})^2 + \sum_{x_i \in R_{1j}^m} (r_{k,i} - C_{2j}^{(k)})^2 \right] \tag{11}$$

In Eq. (11), there are $R_{1j}^{(m)} = \{x_i \mid x_i^{(m)} \le s_j^{(m)}\}, R_{2j}^{(m)} = \{x_i \mid x_i^{(m)} \le s_j^{(m)}\}, C_{kj}^{(m)} = \frac{1}{N_{kj}^{(m)}} \sum_{x_i \in R_{kj}^{(m)}} r_{k,i}$; $N_{kj}^{(m)}$ and $R_{kj}^{(m)}$ denoting the number of samples, the iteration stops when $j = J$ and goes to the fourth step, the optimal output value of k iteration is shown in Eq. (12)

$$C_{kj} = \arg\min_C \sum_{x_i \in R_{kj}} L(y_i, f_{k-1}(x_i) + C) \tag{12}$$

Afterwards, the iterative functions are updated, and the computational equation is shown in Eq. (13).

$$f_k(x) = f_{k-1}(x) + \eta \sum_{j=1}^{J} C_{kj} I(x \in R_{kj}) \tag{13}$$

The function is updated until there is $m = M$; the iteration ends to output the final strong learner $F(x) = f_k(x) = f_0(x) + \eta \sum_{k=1}^{K} \sum_{j=1}^{J} C_{kj} I(x \in R_{kj})$, otherwise the second to fourth steps are repeated. GBDT, as a nonlinear model, can capture more complex data patterns by integrating multiple decision trees, but it is not suitable for parallel computation and processing large-scale datasets. Therefore, the study proposes a model based on the fusion of GBDT and LR, which fully exploits the advantages, disadvantages, and complementarities of both. The research proposed model extends the feature space by multiple leaf nodes generated by GBDT, each representing a discriminative feature or combination of features, and these newly generated features are subsequently used as inputs to the LR model. In this way, the non-linear properties of GBDT and the linear properties of LR can complement each other in the same recommender system, thus achieving higher recommendation accuracy. The model training process is shown in Figure 5.

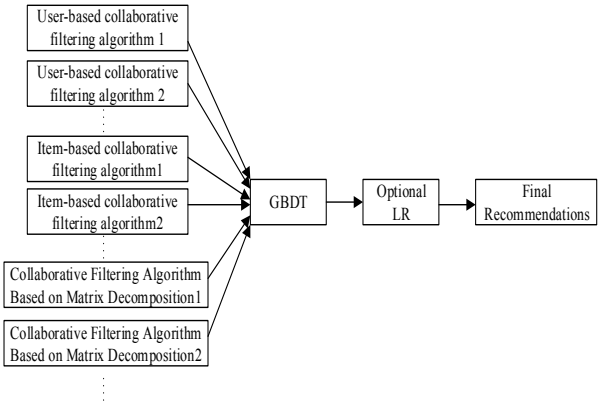

Fig. 5: GBDT-LR model training flow

Before model training, the training data need to be preprocessed. Firstly, by counting the historical transaction data of property companies, the dataset was divided into a training set and a testing set, whose ratio was maintained at 9:1. Subsequently, three different types of collaborative filtering methods were applied to the training set to generate the corresponding prediction models. The collaborative filtering methods are user-based collaborative filtering, house-based collaborative filtering, and matrix decomposition-based collaborative filtering. Eventually, the outputs of these three models were integrated and fed into the GBDT model as a new training set. Next, the output features of the GBDT model are used as inputs to the logistic regression (LR) model. In this step, the GBDT model is first trained to learn the complex relationships and patterns of the training data. Once the GBDT model is trained, its output features (usually the indexes of leaf nodes) are used as input features to the logistic regression model. This was done to take advantage of the non-linear learning capabilities of the GBDT model while utilizing the excellent classification performance of the logistic regression model. The logistic regression model is then trained on these new features to make the final prediction. This combined GBDT and logistic regression approach capitalizes on the strengths of both: the power of GBDT for feature engineering, and the advantages of logistic regression for explanatory and classification accuracy. In this way, the combined GBDT+LR model not only captures nonlinear relationships in the data but also provides strong classification performance, resulting in a highly accurate and interpretable predictive model.

4 Real Estate Marketing Digital Management Model and Real Estate Recommendation Algorithm Performance Testing

After completing the design of the real estate marketing digital system, to verify the performance of the system designed by the research in the actual demand environment, firstly, the system is functionally tested based on the system demand analysis to prove that its function meets the expectations, and then the processing performance of the system on the equipment is tested to prove that its performance processing level meets the requirements of the equipment used for the operation of the system. At the same time, the performance of the real estate recommendation algorithm based on GBDT is tested to evaluate its classification effect.

4.1 Performance Testing of Real Estate Digital Marketing Management System

To ensure that the real estate digital marketing management system can achieve the expected performance and accuracy in practical applications, performance tests are conducted to assess the stability and reliability of each functional module of the system. Firstly, the real estate digital marketing management system is tested for functionality, and the test results are shown in Table 1.

Table 1. Functional test results

Functional Testing	Input data	Expected output	Test results
user login	Wrong username or password	An incorrect username or password prompt appears	Consistent with the expected output
	Correct username and password	Login Successful	Consistent with the expected output
Enquire about property information	Enter property information	Enter the property information to search, only the correct property information can be entered to search, enter the customer's name to search the property will give an error message	Enquiry Successful
	Enter customer name		Prompt for input error
Enquiry of customer information	Customer Name	When you need to query customer information, you need to enter the customer's name to query the fruit query customer information, the customer's birthday entered will prompt an error.	Enquiry Successful
	Customer Birthday		Incorrect client entered
Customer Booking	The customer selects the property and then the salesperson makes the booking	No duplication of listings	Booking Success
		Duplication of salesperson's choice of listings	Please re-select the listing
Customer Payment	Delivery of the property is successful for payment	Customer Payment Successful	as expected

Table 1 shows the results of the functionality testing of the real estate digital marketing management system, covering key functions such as user login, property information query, customer information query, booking, and payment. The results are in line with expectations, indicating that the system's functionality is stable and reliable. Especially in the key operations of user login and property booking, the system can accurately handle various inputs including incorrect credentials and duplicate selections, thus confirming its robustness. Considering modern society, people's extremely high demand for efficiency and the user's equipment performance is not easy, to make the system be universally applied to a wide range of equipment, based on the above functional test results, through the pre-written running script, through the script execution and scenario simulation of the digital platform in use and the CPU occupancy rate of the test, the results are shown in Figure 6.

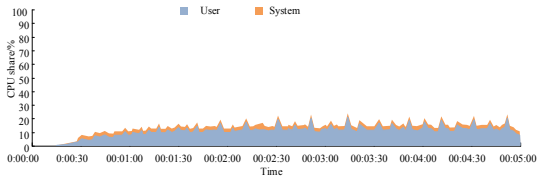

Fig. 6: Performance evaluation results

As shown in Figure 6, it is the result of the system CPU occupancy test during the running time of the script, with the continuous running of the script, the CPU occupancy of the server increases, with an overall upward trend, but with large fluctuations, and the CPU occupancy rises instantly when a new user or a new process joins, after which it drops to the normal level. The CPU occupancy is the highest at 21.7%, and it is the lowest at the start of the process, and stays below 20% during the whole During the entire running period of the script, the CPU usage stays below 20%, which is within the controllable range of server resources. The trend of the system CPU ratio over time is similar to the server resource ratio and slightly higher than the server resource ratio, with the highest CPU ratio of 27.6% during the entire script running time, the lowest CPU occupancy at the beginning of the process, and the CPU occupancy staying below 30% during the entire script running period. Considering the requirements of the digital platform on the real-time nature of the effect, the system CPU occupation ratio should be controlled below 40%, and the test results should be in line with the system design. The results of the response speed test are shown in Figure 7 when the number of multiple accesses.

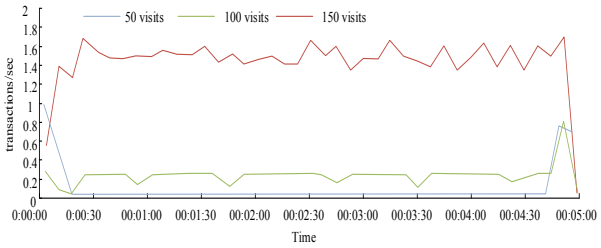

Fig. 7: Response speed test results

In Figure 7, the performance evaluation of the digital platform entry function is carried out for simulated users with different access quantities, and the access quantities are set to 50, 100, and 150, while executing the scripts, and with the change of access quantities, the response speed of the system changes accordingly, and the system's efficiency of the transaction processing is consistent. As can be seen in Figure 7, when the number of concurrent users is 50, the overall change in TPS (Transaction per Second) is smooth over time and maintained at a low level; when the number of concurrent users reaches 100, the fluctuation of TPS is more drastic than that when the number of concurrent users is 50, but it is still smooth; when the number of concurrent users is 150, the fluctuation of TPS is larger, but the execution of the At the same time as the system processing script, the system still maintains good system stability, and its use experience is better, which can meet the use needs of real estate sales staff. The system's task response ability is affected by the system's resources, and the system's response speed will decrease as the number of simultaneous users increases. To assess the actual effectiveness of the recommender system, this study compared the marketing digital management system with the traditional sales management system through the A test. The test was conducted over one month, with each system being used for 15 days each. Considering that weekends and holidays may cause fluctuations in sales, data from these days during the test period were averaged to control for variables. Specific results are shown in Table 2.

Table 2. 15-day comparison test

Evaluation indicators	legacy system	The system proposed by the study
Recommended Satisfaction	82.3	89.7
monthly sales	12.87 million	14.92 million

As shown in Table 2, referral satisfaction is based on customer scores. When using the traditional system, the monthly sales were 12.87 million; after adopting the system of this study, the sales increased to 14.92 million, which is an increase of 15.9 percent. At the same time, this system achieved an increase of 8.9 percent in recommendation satisfaction, indicating that it significantly enhanced customer satisfaction.

4.2 Performance Test of Real Estate Recommendation Algorithm based on GBDT

The study uses datasets derived from customer and property data in the real estate e-marketing system, and divides the training set and test set by 9:1, with 9000 sets of property sales data in the training set and 1000 sets of property sales data in the test set. The collaborative filtering model based on customer information is modeled by six models with six different training parameters, and the main parameters are shown in Table 3.

Table 3. Main parameters of the model

Collaborative filtering algorithm based on user information		Collaborative filtering algorithm based on property information	
K-value	user similarity	K-value	user similarity
K=50	Similarity of Pearson	K=100	Similarity of Pearson
K=50	Euclidean distance similarity	K=100	Euclidean distance similarity
K=300	Similarity of Pearson	K=400	Similarity of Pearson
K=300	Euclidean distance similarity	K=400	Euclidean distance similarity
K=800	Similarity of Pearson	K=900	Similarity of Pearson
K=800	Euclidean distance similarity	K=900	Euclidean distance similarity

As shown in Table 3, the number of iterations of the matrix decomposition-based recommendation algorithm is 100, 800, and 1500 cases, and the models obtained from the above three algorithms are synthesized as inputs to the GBDT algorithm. Evaluating model performance is a crucial step in machine learning and data analytics. ROC (Receiver Operating Characteristic) curve is a commonly used evaluation tool, especially when dealing with classification problems. By comparing the ROC curves of different classifiers, it is possible to visualize the performance of each model in terms of true-positive and false-positive rates, so that the optimal model can be more accurately selected. A comparison of the ROC curves of multiple

classifiers in a real estate digital marketing management system is shown in Figure 8.

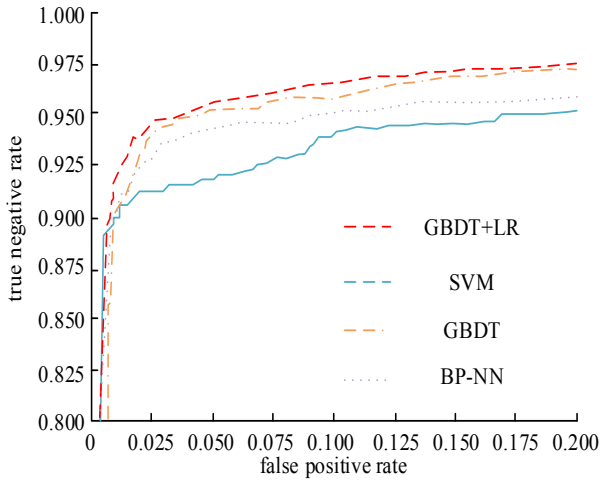

Fig. 8: ROC curve analysis

In ROC curve analysis, the performance of a classifier is usually measured by the area under the curve. According to the data in Figure 8, the combined model of GBDT plus logistic regression achieved 92% in training accuracy, which is significantly better than the 89% of the single GBDT model, 83% of the BP neural network, and 69% of the SVM. These results indicate that the GBDT plus logistic regression model performs the best of all the classifiers compared and is therefore most suitable for use as a performance classification evaluator. This further emphasizes the superiority of the combined GBDT and logistic regression model in terms of classification accuracy. Using the model parameters shown in Table 1, the user collaborative filtering model, the commercial property collaborative filtering model, the collaborative filtering model with matrix decomposition, the linear weighting algorithm, the fusion hybrid recommendation algorithm, and the random forest algorithm are used as experimental comparisons to compare the recall and accuracy of the different recommendation models, and the experimental results are shown in Table 4.

As shown in Table 4, the hybrid GBDT model performs well in terms of Precision, Recall, Accuracy, and F1 scores, with an accuracy of 94.63% and a regression rate of 94.82%, when compared to the pure user collaborative filtering model with other hybrid recommendation algorithms. This is because GBDT itself is a strong classifier that can handle complex non-linear relationships, but when there are linear relationships in the data, GBDT may not perform as well as a linear model.

Table 4. Comparative results of recall and accuracy of models

Classification algorithm	Precision	Recall	Accuracy	F1
Hybrid bit	0.9445	0.9482	0.9463	0.9463
Collaborative user filtering	0.9123	0.9123	0.9100	0.9123
Commodity co-filtering	0.8935	0.8921	0.8900	0.8928
Matrix decomposition	0.8279	0.8757	0.8500	0.8513
Linear weighting	0.9013	0.9003	0.8990	0.9008
Fusion hybrid recommendations	0.8839	0.8897	0.8850	0.8868
Random forest	0.9219	0.8961	0.9050	0.9089

By combining GBDT with LR, the model not only captures the nonlinear characteristics of the data but also accurately models the linear relationship of the data, thus achieving higher classification accuracy. Recommendation accuracy is divided into Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), the smaller the value of MAE and RMSE, the higher the recommendation accuracy. The comparative test results of RMSE and MAE values are shown in Figure 9.

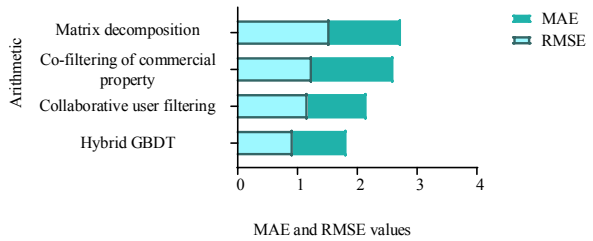

Fig. 9: Comparative results of RMSE and MAE values

As can be seen in Figure 9, it can be seen that the hybrid GBDT algorithm used in the study also has lower RMSE and MAE values than most of the collaborative filtering recommendations based on classical similarity metrics, and the results in RMSE are improved by 22.1 percent over the sub-optimal algorithm (User Collaborative Filtering), and the results of the hybrid GBDT proposed in the study are 10.1 percent better than the sub-optimal algorithm (User Collaborative Filtering) in terms of MAE.

5 Conclusion

In the context of the current digital economy's rapid development, the real estate industry is also gradually shifting towards smarter and more automated management systems. The study proposes a real estate digital marketing management system applicable to the real estate marketing field. The performance test results of the digital marketing management system show that the system is consistent with the expected outputs in terms of key functional modules such as user login, property query, customer information query, property booking, and payment, the CPU occupancy rate basically stays below 30% during the whole script running period, and the system still maintains a better system stability when the concurrent amount of TPS users is 150, which makes its usage experience better. The performance test results of the GBDT-LR model show that its accuracy is as high as 94.63%, and the regression rate is 94.82%, which is particularly good in terms of classification accuracy, and the comparison of the ROC curves shows that compared with the 89% of the single GBDT model, the 83% of the BP neural network, and the 69% of the SVM, the accuracy of the GBDT-LR model is as high as 92%. The experimental results show that the GBDT-LR model can model the linear relationship of the data with certainty and achieve higher classification accuracy. However, there are still limitations in the study and although the hybrid GBDT model performs well in terms of accuracy, the model structure needs to be optimized and adapted when dealing with large-scale data to improve efficiency.

References:

[1] Mohamed OA, Ahmed KM, Peter M, Shehab EIB, Minatallah MH, Mohamed MA, Ahmed A, Mohamad HE, Ashraf N, Sameh OA, Optimizing the artificial lighting in a smart and green glass building-integrated semi-transparent photovoltaics: a multifaceted case study in Egypt, *WSEAS Transactions on Environment and Development*, Vol. 17, 2021, pp. 118-127, https://doi.org/10.37394/232015.2021.17.12.

[2] Ho HC, Song Y, Cheng W, Liu Y, Guo Y, Lu S, Lum T, Har Chiu RL, Webster C, How do forms and characteristics of Asian public housing neighbourhoods affect dementia risk among the senior population? A cross-sectional study in Hong Kong, *Public Health*, Vol. 219, 2023, pp. 44-52.

[3] Tomczak M, Jaśkowski P, Harmonising construction processes in repetitive construction projects with multiple buildings, *Automation in Construction*, Vol. 139, 2022, pp. 104266.

[4] Vera M, Berescu C, Macri Z, An uncertain future: prospects for Bucharest's large housing estates, *Journal of Housing and the Built Environment*, Vol. 38, No. 1, 2023, pp. 101-119.

[5] Chmielewska A, Ciski M, Renigier-Bilozor M, Residential real estate investors' motives under pandemic conditions, *Cities*, Vol. 128, 2022, pp. 103801.

[6] Yan PC, Sun QS, Yin NN, Hua LL, Shang SH, Zhang CY, Detection of coal and gangue based on improved YOLOv5.1 which embedded scSE module, *Measurement*, Vol. 22, No. 3, 2022, pp. 530-542.

[7] Han S, Lei Z, Hermann U (Rick), Bouferguene A, Al-Hussein M, 4D-based automation of heavy lift planning in industrial construction projects, *Canadian Journal of Civil Engineering*, Vol. 48, No. 9, 2021, pp. 1115-1129.

[8] Jia W, Xu SQ, Liang Z, Zhao Y, Min H, Li SJ, Yu Y, Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector, *IET Image Processing*, Vol. 15, No. 14, 2021, pp. 3623-3637.

[9] Chowdhury AM, Moon S, Generating integrated bill of materials using mask R-CNN artificial intelligence model, *Automation in Construction*, Vol. 145, 2023, pp. 644-658.

[10] Cao L, Zhang W, Kan X, Yao W, A novel adaptive mutation PSO optimized SVM algorithm for sEMG-based gesture recognition, *Scientific Programming*, Vol. 2021, 2021, pp. 9988823.

[11] Ezaldeen H, Bisoy SK, Misra R, Alatrash R, Semantics-aware context-based learner modeling using normalized PSO for Ted E-learning, *Journal of Web Engineering*, Vol. 21, No. 4, 2022, pp. 1187-1223.

[12] Rajani B, Sekhar DC, A hybrid optimization based energy management between an electric vehicle and electricity distribution system, *International Transactions on Electrical Energy Systems*, Vol. 31, No. 6, 2021, pp. 905-935.

[13] He WH, Meng H, Han J, Zhou GH, Zheng H, Zhang SL, Spatiotemporal PM2.5 estimations in China from 2015 to 2020 using an

improved gradient boosting decision tree, *Chemosphere*, Vol. 296, 2022, pp. 134-145.

[14] Lemus Cardenas L, Astudillo L, Juan P, Mezher A, GraTree: a gradient boosting decision tree based multimetric routing protocol for vehicular ad hoc networks, *Ad Hoc Networks*, Vol. 137, 2022. pp. 106-117.

[15] Zhang ZQ, Li L, Li X, Hu YC, Huang K, Xue BY, Wang YQ, Yu YJ, State-of-health estimation for the lithium-ion battery based on gradient boosting decision tree with autonomous selection of excellent features, *International Journal of Energy Research*, Vol. 46, No. 2, 2022, pp. 1756-1765.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Shuangxin Chen carried out the research concept and design and writing the article.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The author has no conflicts of interest to declare.