Research Article

Research on Precision Marketing of Real Estate Market Based on Data Mining

Rong Huang and Shuai Mao

Department of Economics and Management, Hubei University of Automotive Technology, Shiyan 442000, Hubei, China

Correspondence should be addressed to Shuai Mao; 20120014@huat.edu.cn

Received 22 January 2022; Revised 15 February 2022; Accepted 22 February 2022; Published 28 March 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Rong Huang and Shuai Mao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As one of the basic industries in China, the real estate industry contributes much to the national GDP every year and plays an important role in stimulating the economy. After years of development, the real estate industry has accumulated a large number of sales data, includes the data of customer and construction. However, the utilization of them is still in the stage of extensive collection and use, which is not sufficient for the accumulated data. Therefore, a precise marketing management system for real estate enterprises based on data mining technology is developed in this paper. Through big data mining, the target customers can be accurately subdivided and portrayed, which realizes the matching and prediction of resources. On this basis, accurate marketing and promotion of multichannel collaboration are implemented, which realizes the innovation of real estate marketing mode in the era of big data.

1. Introduction

With the rapid development of information technology, the use of Internet+, 5G, cloud computing and mobile terminals has brought us into the era of big data [1]. Big data technology has become a hot topic at present that the arrival of big data era promotes the interaction between human beings and makes exchange of information more convenient, which also makes the Internet economy more closely linked, so that people can create wealth without leaving home [2–4]. “Big data” has penetrated into all walks of life, which subverts the operation thinking and marketing mode of many traditional industries. In this context, big data touches the nerves of the real estate industry, and real estate enterprises can accurately understand the demand of consumers, so as to formulate accurate marketing strategy [5].

As an emerging field, data mining has a wide range of application prospects, and has been widely used in all walks of life [6, 7]. The combination of real estate field and data mining technology is a topic that researchers are devoted to studying in recent years. It is meaningful to using data mining technology to research and analyze the application of big data in real estate marketing.

2. Overview of Data Mining Technology

Precision marketing is based on the collation and analysis of customer’s data, which can accurately grasp the demand of customers, provide customers with appropriate products and marketing means, and realize the company’s interests. Data mining, including clustering analysis, discriminant analysis and factor analysis, is commonly used in practical technology [8].

2.1. Definition of Data Mining. Berry and Linoff [9] defined data mining as the process of exploring and analyzing large amounts of data in order to discover meaningful patterns and rules. A large number of data may be partly noise data or fuzzy data. The object of data mining can be a database, a file system, or any other data collection organized together.

The data set in the real estate database includes real estate sales data and feature data of real estate. By randomly extracting part of the data, and data conversion, analysis and other processing, we can find the key data that needed for the development of real estate sales strategy [10, 11].
Data mining is divided into directed one and non-oriented. The purpose of directed data mining is to interpret or classify a specific target domain. The purpose of non-directional data mining is to find out the pattern or similarity between batch data without presetting target domain or class.

2.3. Methods of Data Mining. The purpose of data mining is to mine valuable information. In order to realize data mining, we need to adopt certain methods, which is helpful to the better realization of data mining [18]. It can be completed by different methods and means, while users can choose the appropriate method according to their own needs. The following content describes several common methods:

2.3.1. Data Summary. The word "summary" is not difficult to understand literally, which is to use simple and concise sentences to summarize the more complex issues, so that people can understand the content of the exposition in a short time. In the same way, data summary is to concentrate the existing data in certain ways, such as through statistical methods such as sum and average. The calculation results of the data are reflected by charts. In addition, the common chart models are column chart and pie chart [19]. Data mining is actually a process of data summary, but its research is more in-depth, which needs a more comprehensive summary of the data from a deep level and a wide angle. Generally, the data are analyzed, synthesized, abstracted and summarized in order from low level to high level, so as to find out some internal relations of these data, and judge the direction of future data development through these laws. According to the methods, it can be divided into multidimensional data analysis method and attribute oriented induction method.

2.3.2. Classification Mining. Classification mining refers to the classification of data in order to mine its potential value. It mainly classifies the data by mapping, so as to correctly classify the data to be processed. There are many ways to construct classifiers, while statistical method, machine learning method and decision tree method are commonly used [20, 21].

2.3.3. Cluster. Cluster refers to aggregate classification, that is, the data with the same characteristics are classified, aggregated and stored separately. In this way, the relationship between data is very clear, and the data belonging to the same category have the same or similar categories; on the contrary, data in different categories have different categories which is helpful to the later work. Statistical method and machine learning method are commonly used in clustering methods.

3. Design of Precise Marketing Management System Based on Data Mining

3.1. Demand Analysis. The purpose of functional requirements analysis is to clarify the functional indicators of the real estate precise marketing management system, that is, the function points that the system should have. In the development of software, there are many tools used to describe users' requirements. This section will discuss them in detail through use case diagrams.

The overall use case of precision marketing management system is shown in Figure 2. The users of the target system mainly include enterprise leaders, market leaders, salesmen and system administrators. In addition, the business of the system can be roughly divided into real estate management, house management, building management, sales management, customer management, decision support and system management.

The real estate management mainly completes the maintenance of the basic information of the real estate, including the business of adding, deleting, modifying and querying the information of the real estate, and the users mainly include the person in charge of the market and the system administrator.

Building management mainly realizes the maintenance of building basic information, including the business of building information addition, building information deletion, building information modification, building information query and so on. The users involved include market leader and system administrator.

Housing management is the management of room information, which mainly maintains the specific information of the room, including the business of adding housing information, deleting housing information, modifying
housing information, querying housing information, etc., and the users involved are market leaders and system administrators.

Sales management mainly maintains the basic information of real estate sales, including sales opportunity management, sales record management, sales performance management, etc., and the users involved are market leaders and salesmen.

Customer management includes customer information addition, customer information deletion, customer information modification, customer information query, etc. the user roles involved include enterprise leader, market leader and salesman. Among them, enterprise leader can only be responsible for customer information query.

Decision support includes sales forecasting, performance statistics, weekly and monthly performance reports. The users involved include enterprise leaders, market leaders and salesmen.

System management is mainly for system administrators, providing the maintenance of basic system information, including user management, data backup, data restore, permission setting, etc.
3.2. Overall Framework of System. In the overall design of the system, first of all, we need to design the overall framework of the system. The target system is mainly used for the daily sales management of real estate enterprises which mainly targets at the enterprise leaders, market leaders, salesmen and system administrators, and provides different services for diverse users. According to the actual needs of the system, B/S mode is selected as the structure, which is a distributed application architecture based on Web. It can meet the needs of different types of users and is suitable for the development of precision marketing management system [23].

In the process of execution, it will give all the interaction between users to the browser side of the system, that is, the interface layer, and the business and data related operations are handed over to the web server to complete, so that the browser and the web server work together to complete the processing of requests. The architecture of precision marketing management system is shown in Figure 3.

3.2.1. Interface Layer. The interface layer is usually called user layer or application layer. The main function of this layer is to complete the interaction with users. On the one hand, it receives the request messages sent by users to the system, on the other hand, it feeds back the request results completed by the server to users for browsing. For users, this layer is the intuitive that their evaluation of the software will be directly reflected in the user’s sense of operation on the interface layer. When a user sends out an access request, the interface layer will receive the request, and then send it to the business logic layer and data access layer. At the same time, the processing results will be fed back to the user’s browser through HTML and displayed to the user.

3.2.2. Business Logic Layer. The business logic layer is responsible for executing the part that needs business logic judgment and processing in the user request which is located in the middle of the three layers. It is a bridge between the client and the database and it can be said that it is the core layer of the three layers, and its role is very important. When the interface layer receives the user’s request, it will send the request of the logical processing part to the layer, which is responsible for executing it. At the same time, it will also send the request of data processing to the data access layer.

In the three layers, the business logic layer has its own responsibilities from the top to the bottom. It also acts as a callee’s identity, in this layer, involves a lot of business relations. Therefore, the business logic layer is the core part of the whole architecture.

3.2.3. Data Access Layer. Data access layer is also known as application data source layer. The main function of this layer is to complete the operation of database, including data call and data processing. Through the access layer, you can query, modify, delete and update database tables, and it will provide data call and execution related services for the middle layer. Because the operation of the system cannot be separated from the operation, and the database design of any system is relatively complex and takes up more resources, which makes the performance requirements of the database system higher, so the access mode of the database system should be optimized as much as possible, in order to improve the overall efficiency of the system.

3.2.4. Advantages of Three-Tier Architecture. The reason why the architecture design adopts the three-tier architecture mode is as follows:

(i) The three-tier partition makes the system more flexible, and realizes the maintainability and expansibility of the later functions and performance. Because of the independence of the three layers, it is easy to transplant the database

(ii) It accords with the design idea of “high cohesion, low coupling” in software engineering

(iii) The three layers are independent of each other, and the related functions are relatively clear, which facilitates the development of developers, effectively improves the efficiency of system development work, and shortens the cycle

3.3. Deployment and Functional Structure of System. The precision marketing management system based on Net environment is discussed in this paper. B/S mode is adopted for system structure. C# is used for foreground programming, SQL Server 2005 is used for backstage database, and ADO is applied between foreground and background Net connection to achieve interaction. For the system developed by B/S mode, the deployment can be divided into three parts: client, application server, and database server. The request submitted by the client will be sent to the application server for processing, and then through the operation of the database, the response to the user’s request is realized. The deployment of the system is shown in Figure 4.

In the work of system design, the first is to design the overall architecture. When the overall architecture is determined, the design will have an overall direction. The next work is to design the function of the system, that is, to design some functions that the target system needs. This is the most concerned problem of users, which means, it can help users complete specific functions. The target system is applied in the real estate sales management. Therefore, in the demand analysis stage, the author talked with the staff related to sales management in depth, and defined the functional requirements of the system, which also laid a good foundation for the design of functional modules. The requirements of modular design are summarized as follows:

(1) In the process of module division, the system should be divided according to the hierarchy, that is, the system should be first divided from the overall perspective, and then the modules after division should be further divided, so that the final module can be realized in a certain way.
All the modules divided should be as independent as possible. In other words, there should be no association between these modules under special circumstances. Of course, this situation is not absolute, but should be avoided as far as possible.

The relationship between the divided modules must be clear. The result of this is to facilitate the later maintenance of the system. When facing problems, the programmer can easily modify it and trace it. After understanding the requirements of the system function module division, we can divide the function modules of the precise marketing management system. The relevant functional modules mainly include eight functional modules, including real
3.4. Design of Database. According to the process of concept design is the process of data entity design. Based on the previous analysis of the real estate precision marketing management system, part of the data entity design of the system is given below. The main entities of the system include user entity, house entity, building entity, room entity, customer entity, sales contract entity, etc.

The content of database logical design is based on the entities obtained from conceptual design, and transforms each entity into the actual data physical storage structure. This section gives the relevant data tables of the real estate precision marketing management system, and each data table includes the attribute description of each entity. The data tables of the target system include user information table, house type table, room information table, building information table, property information table, property sales information table, customer information table, sales contract table, etc.

The structure of user information table is shown in Table 1. The data table is used to store the basic user information of the system, including the fields of user ID, user name, user type, user real name, gender, age, etc., in which the user ID is the primary key.

The structure of the house type table is shown in Table 2. The data table is used to store the type information of the house. While this table includes mainly house type ID, house type name and description, in which the house type ID is the primary key.

The structure of the room is shown in Table 3. The data table is used to store the basic information of the room, including the fields of room ID, room number, floor area, building ID, unit number, and unit ID, etc. Among them, the room ID is the primary key, and the building ID and unit type ID are foreign keys.

The structure of the building is shown in Table 4. The data table is used to store the building type information, including the fields of building type ID, type name, type description, etc., in which the building type ID is the primary key.

The structure of the building includes the building ID and unit type ID, which are foreign keys.

The structure of the real estate is shown in Table 5. The data table is used to store the basic information of the real estate. The fields mainly include the property name, developer, floor area, building area, etc., in which the real estate ID is the primary key of the data table.

The structure of the real estate sales information table is shown in Table 7. The data table is used to store the sales of real estate. The main fields include sales ID, ID, ID, average price, and number of households sold. Among them, the sales ID is the primary key, and the real estate ID, building ID and room ID are all foreign keys.

The structure of the customer information is shown in Table 8. The data table is used to store the basic information of customers, including the fields of customer ID, customer name, contact number, age, occupation, etc., in which the customer ID is the primary key of the data table.

The structure of the sales contract table is shown in Table 9. The data table is used to store the sales contract information. The fields include sales contract ID, customer name, purchase time, purchase price, payment method, etc. Among them, the sales contract ID is the primary key of the data table.

3.5. Design of Interface and Operation

3.5.1. Interface Design. The interface design mainly includes the definition of external interface and internal interface. The detailed design is shown in Table 10.
### Table 1: Table of users’ information.

| Field name                | Field type | Data length | Primary key/foreign key | Allow to be empty |
|---------------------------|------------|-------------|-------------------------|-------------------|
| User ID                   | Int        | 32          | Primary key             | No                |
| Username                  | Varchar    | 20          | No                      | No                |
| User type                 | Varchar    | 10          | No                      | No                |
| User’s real name          | Varchar    | 50          | No                      | Yes               |
| Gender                    | Byte       | 1           | No                      | Yes               |
| Age                       | Int        | 11          | No                      | Yes               |
| User password             | Varchar    | 50          | No                      | No                |
| E-mail                    | Varchar    | 50          | No                      | Yes               |
| Mobile phone              | Varchar    | 20          | No                      | Yes               |
| User information remarks  | Varchar    | 255         | No                      | Yes               |
| Last login time           | Datetime   | 8           | No                      | No                |
| User’s current state      | Int        | 32          | No                      | No                |

### Table 2: House type.

| Field name                | Field type | Data length | Primary key/foreign key | Allow to be empty |
|---------------------------|------------|-------------|-------------------------|-------------------|
| Unit ID                   | Int        | 32          | Primary key             | No                |
| House name                | Nvarchar   | 50          | No                      | No                |
| House description         | Nvarchar   | 50          | No                      | Yes               |

### Table 3: Room information table.

| Field name                | Field type | Data length | Primary key/foreign key | Allow to be empty |
|---------------------------|------------|-------------|-------------------------|-------------------|
| Room ID                   | Int        | 32          | Primary key             | No                |
| Building ID               | Int        | 32          | Foreign key             | No                |
| Floor                     | Nvarchar   | 3           | No                      | No                |
| Room No                   | Nvarchar   | 50          | No                      | No                |
| Unit ID                   | Int        | 32          | Foreign key             | No                |
| Use                       | Nvarchar   | 50          | No                      | Yes               |
| Area                      | Int        | 50          | No                      | No                |
| Construction area         | Float      | 7           | No                      | No                |
| Starting price            | Money      | 7           | No                      | No                |
| Highest price             | Money      | 7           | No                      | No                |
| Average selling price     | Money      | 7           | No                      | No                |
| Sales discount            | Nvarchar   | 4           | No                      | No                |
| Selling price             | Money      | 7           | No                      | No                |
| Balcony area              | Float      | 7           | No                      | No                |
| Room orientation          | Nvarchar   | 50          | No                      | No                |
| Decoration standard       | Nvarchar   | 500         | No                      | Yes               |
### Table 4: Information of building types.

| Field name         | Field type | Data length | Primary key/foreign key | Allow to be empty |
|--------------------|------------|-------------|-------------------------|-------------------|
| Type ID            | Int        | 32          | Primary key             | No                |
| Type name          | Nvarchar   | 50          | No                      | No                |
| Type description   | Nvarchar   | 50          | No                      | Yes               |

### Table 5: Information of building.

| Field name           | Field type | Data length | Primary key/foreign key | Allow to be empty |
|----------------------|------------|-------------|-------------------------|-------------------|
| Building ID          | Int        | 32          | Primary key             | No                |
| Property ID          | Int        | 32          | Foreign key             | No                |
| unit number          | Nvarchar   | 10          | No                      | No                |
| Area                 | Float      | 7           | No                      | No                |
| Construction area    | Float      | 7           | No                      | No                |
| Building type ID     | Int        | 32          | Foreign key             | No                |
| Number of layers     | Varchar    | 10          | No                      | No                |
| Number of garages    | Int        | 32          | No                      | No                |
| Number of households per floor | Int | 32 | No | No |
| Consumer groups      | Nvarchar   | 30          | No                      | No                |
| Remark               | Text       | 250         | No                      | Yes               |
| Total houses         | Int        | 32          | No                      | Yes               |
| Sold                 | Int        | 32          | No                      | Yes               |
| Unsold               | Int        | 32          | No                      | Yes               |

### Table 6: Information of property.

| Field name             | Field type | Data length | Primary key/foreign key | Allow to be empty |
|------------------------|------------|-------------|-------------------------|-------------------|
| Property ID            | Int        | 32          | Primary key             | No                |
| Property name          | Varchar    | 30          | No                      | No                |
| Developer              | Varchar    | 50          | No                      | No                |
| Design unit            | Varchar    | 30          | No                      | No                |
| Regional lot           | Varchar    | 20          | No                      | No                |
| Architectural style    | Varchar    | 20          | No                      | No                |
| Area                   | Float      | 7           | No                      | No                |
| Construction area      | Float      | 7           | No                      | No                |
| Total number of rooms  | Int        | 35          | No                      | No                |
| Price method           | Varchar    | 10          | No                      | No                |
| Development time       | Datetime   | 8           | No                      | No                |
| Opening time           | Datetime   | 8           | No                      | No                |
| Opening method         | Varchar    | 50          | No                      | No                |
| Traffic condition      | Varchar    | 100         | No                      | Yes               |
| Property management    | Varchar    | 100         | No                      | Yes               |
| Remark                 | Text       | 250         | No                      | Yes               |
| File creation date     | Datetime   | 8           | No                      | Yes               |
| Sales call             | Varchar    | 20          | No                      | Yes               |

### Table 7: Information of property sales.

| Field name              | Field type | Data length | Primary key/foreign key | Allow to be empty |
|-------------------------|------------|-------------|-------------------------|-------------------|
| Property sales ID       | Int        | 32          | Primary key             | No                |
| Property ID             | Int        | 32          | Foreign key             | No                |
| Building ID             | Int        | 32          | Foreign key             | No                |
| Room ID                 | Int        | 32          | Foreign key             | No                |
| Average selling price   | Float      | 7           | No                      | No                |
| Number of households sold | Int    | 32          | No                      | No                |
| Sales rate              | Varchar    | 20          | No                      | Yes               |
| Lowest price            | Money      | 7           | No                      | Yes               |
| Highest price           | Money      | 7           | No                      | Yes               |
| Discount                | Float      | 7           | No                      | Yes               |
| Promotion channels      | Nvarchar   | 50          | No                      | Yes               |
| Total houses            | Int        | 32          | No                      | Yes               |
| Predetermined quantity  | Int        | 32          | No                      | No                |
| Unsold quantity         | Int        | 32          | No                      | No                |
### 3.5.2. Operation Design

1. **Combination of Running Modules.** The system mainly takes a window as a module. Generally, a window completes a specific function. While the main window realizes the connection and combination of different functions between modules by opening another sub window. Each module is relatively independent, and the program has good portability. Moreover, the cooperation and data sharing between modules are realized by transferring the reference of data items.

2. **Operation Control.** The user opens the system login window and enters the name and password. Then the system jumps to the corresponding background according to the user type corresponding to the name, so as to realize different operations with diverse permissions and roles.

3. **Running Time.** The running time of each module should be controlled within 1–2 seconds (most of which is in response to the user’s action). As the system adopts message driven mode, it will effectively improve the utilization of computer.

### 3.6. Algorithm of Decision Tree Construction and ID3 Learning

#### 3.6.1. Decision Tree Construction Algorithm.**

The construction algorithm of decision tree can be completed by training set \( T \), where \( T = \{ \langle x, C_i \rangle \} \), and \( x = (a_1, a_2, \ldots, a_n) \) is a training example, it has \( n \) attributes listed in the attribute table \( \langle A_1, A_2, \ldots, A_n \rangle \) where \( a_i \) is the value of attribute \( A_i \). \( C_i \in C = \{ C_1, C_2, \ldots, C_m \} \) is the classification result of \( X \) [24–26]. The algorithm is divided into the following steps:

1. Select the attribute \( A_1 \) from the attribute table as the classification attribute; If there are \( K_i \) values in attribute \( A_i \), \( T \) is divided into \( k_i \) subsets \( T_{ij} \), \( i \leq K_i \), where

   \[
   T_{ij} = \{ \langle x, C_i \rangle \mid \langle x, C_i \rangle \} \in T \text{ and the attribute value } a \text{ of } X \text{ is the } K_i \text{ value;}
   \]

2. Delete the attribute \( A_i \) from the attribute table

3. For each \( T_{ij} (1 \leq j \leq K_i) \), order \( T = T_{ij} \)

4. If the property sheet is not empty, return (1), otherwise output

   At present, the mature decision tree methods are ID3 and C4.5. Cart, SLIQ, etc.

### Table 8: Management table of customer information.

| Field name               | Field type | Data length | Primary key/foreign key | Allow to be empty |
|--------------------------|------------|-------------|-------------------------|-------------------|
| Customer ID              | Int        | 32          | Primary key             | No                |
| client’s name            | Varcha     | 60          | No                      | No                |
| Employer                 | Varcha     | 60          | No                      | No                |
| ID number                | Varcha     | 20          | No                      | No                |
| Gender                   | Varcha     | 2           | No                      | No                |
| Contact number           | Varcha     | 20          | No                      | No                |
| Address                  | Varcha     | 150         | No                      | No                |
| Age                      | Int        | 32          | Yes                     | Yes               |
| Profession               | Varcha     | 50          | Yes                     | Yes               |
| Demand area              | Float      | 7           | Yes                     | Yes               |
| Demand floor             | Float      | 7           | Yes                     | Yes               |
| Demand room type         | Varcha     | 50          | Yes                     | Yes               |
| Payment method           | Varcha     | 50          | Yes                     | Yes               |
| Purchase intention       | Varcha     | 100         | Yes                     | Yes               |
| Number of homes purchased| Int        | 32          | Yes                     | Yes               |
| Remark                   | Varcha     | 250         | Yes                     | Yes               |

### Table 9: Contract table of sales.

| Field name               | Field type | Data length | Primary key/foreign key | Allow to be empty |
|--------------------------|------------|-------------|-------------------------|-------------------|
| Sales contract ID        | Int        | 32          | Primary key             | No                |
| Customer name            | Varchar    | 20          | No                      | No                |
| Purchased room           | Varchar    | 20          | No                      | No                |
| Price                    | Float      | 7           | No                      | No                |
| Payment method           | Varchar    | 50          | No                      | No                |
| Salesperson              | Varchar    | 50          | No                      | No                |
| Sales manager            | Varchar    | 50          | No                      | No                |
| Financial officer        | Varchar    | 50          | No                      | No                |
| Payment record           | Varchar    | 250         | No                      | No                |
| Related procedures       | Varchar    | 250         | No                      | No                |
| Delivery time            | Datatime   | 8           | No                      | No                |
| Contract remarks         | Text       | 250         | No                      | Yes               |
| Signing time             | Datatime   | 8           | No                      | No                |
| Signatory                | Varchar    | 20          | No                      | No                |
3.6.2. ID3 Learning Algorithm. Information entropy is called average information quantity in information theory, which is an average value used to measure the information transmitted, which includes a finite number of mutually exclusive and joint complete events. They all appear with a certain probability, which is represented by the mathematical formula [27]: a group of events $X_1, \cdots, X_r$ appears with a given probability $p(X_1), \cdots, p(X_r)$ while the mean value $H(x)$ is the information entropy, and its value is equal to the mathematical expectation of the (self) information quantity $I(x)$ of each event

$$H(X) = - \sum_{r=1}^{r} p(X_i)I(X_i) = - \sum_{r=1}^{r} p(X_i)\log p(X_i).$$

(1)

In the traditional ID3 algorithm, the information entropy is used as the standard of attributes selection, and the value of information entropy is obtained based on data calculation. Then it is selected by comparing the size of each information entropy, and the item corresponding to the information entropy is taken as the root node of the decision tree. After the example set is divided into subsets by using this attribute, the entropy value of the system is the minimum. It is expected that the average path of the nonleaf node to reach each descendant leaf node is the shortest, and the average depth of the decision tree generated is smaller [28]. In addition, it can be seen that the more fuzzy and disorderly the training case set is in target classification, the higher its entropy is, the clearer the training case set is in target classification, while the more ordered it is, and the lower its entropy is. ID3 algorithm is based on the principle of “the attribute with greater information gain is more beneficial to the classification of training cases”. In each step of the algorithm, “the attribute in the table that can best classify the training case set” is selected. Moreover, the information gain of an attribute is the decrease of system entropy due to the use of this attribute to divide the sample. The key operation of ID3 algorithm is to calculate and compare the information of each attribute [29, 30].

The above detailed introduction of ID3 algorithm, in order to better achieve data mining, here will be the basic strategy of ID3 algorithm. The implementation of ID3 algorithm is as follows:

1. Each node given in the training sample is taken as the root node of the decision tree to start the process of creating the decision tree.
2. These root nodes are judged and analyzed. If they belong to the same class, they are set as leaf nodes, and the nodes set as leaf nodes are marked.
3. For the samples that do not belong to the same class, the entropy based measure that called information gain is used as the heuristic information, and the best attribute that can be reclassified is selected from the heuristic information, which becomes the test or decision attribute of the node.
4. Create a new branch for the new test property and divide new samples accordingly.
5. The recursive method is used to create a decision tree for each sample. At this time, the attributes of its descendants can not be considered too much.
6. The recursive process should be repeated until the following features appear:
   i. All the attributes represented by a node belong to the same category.
   ii. The attribute represented by the node does not have the feature of continuous partition.
   iii. If there are no samples under the branches of a tree, the classes in these training sets can be compared and analyzed, and the sample of the largest class’s genus is set as the leaf node of the decision tree.

4. System Implementation and Testing

4.1. Implementation of Data Mining

4.1.1. Decision Tree Generation and ID3 Algorithm. The ID3 algorithm proposed by J. R. Quinlan is an earlier and most famous decision tree induction algorithm. Given a set of nonclass attributes $C_1, C_2, \ldots, C_m$, Category attribute $C$ and record training set $S$, a decision tree can be constructed by ID3 algorithm. The ID3 algorithm of decision tree induction algorithm is described as follows [31].

//Returns a decision tree
Function ID3 (R: a nonclass attribute set, C: a Category attribute, s: a training set)
Begin
If $s$ is null, a single node with the value of failure is returned; If $s$ is composed of records whose values are the same category attribute values, and returns a single node with this value; If $R$ is null, a single node is returned, whose value is the most frequent Category attribute value found in $s$ record; The attribute with the maximum gain ($D, s$) value between attributes in $R$ is assigned to $d$; Assign the value of attribute $d$ to $\{d_{j} | j = 1, 2, 3, \ldots, m\}$; The subsets of $s$ composed of records corresponding to $d$, corresponding to $D$ are assigned to $\{s_{1}, 2, 3, \ldots, m\}$; Then return a tree whose root is marked $D$ and its branch is marked $d_{1}, d_{2}, d_{3}, \ldots, d_{m}$;
Then the following trees are constructed
ID3(R-{D}, C, S1), ID3(R-{D}, C, S2), ..., ID3(R-{D}, C, Sm);
End ID3;

4.1.2. Application of Data Mining in Marketing Management.
Combined with the previous analysis, this section analyzes the implementation of data mining with specific cases. Suppose that the following customer information exists in the database, as shown in Table 11.

(1) Step 1: transformation of data. According to the basic data of customer, the required data is transformed by generalization to higher-level concepts which are given as follows:
Data by age are shown in Table 12.
Statistics by income are shown in Table 13.
According to the statistics of purchase area, the statistics are shown in Table 14.
Statistics by marital status are shown in Table 15.

(2) The second step. Get the expected information and information gain. The key to construct a good decision tree is how to choose good logical judgment or attribute. It has been found that the smaller the tree is, the stronger the prediction ability is. To construct a decision tree as small as possible, the key is to choose the appropriate logical judgment or attribute. Information gain here is used to select attributes. The calculation formula of the degree is as follows
Among them, the data set is S, M is the classification number of S, CI is a certain classification label, PI is the probability that any sample belongs to CI, and Si is the number of samples on classification CI.

(1) Entropy divided by a into subsets: A is an attribute with V different values;

(2) Information gain: Gain (a) = I (S1, S2, ..., Sm) – E (A)
The split attribute is obtained as follows:
For purchase area:
Info (purchase area) = 0.999215879
Age:
Expected information: info (age) = 0.965202313
Information gain: gain = 0.034013566
Gender:
Expected information: info (gender) = 0.999202348
Information gain: gain = 0.000013531
Occupation:
Expected information: Info = 0.717207246
Information gain: gain = 0.282008633
Marriage or not:
Expected information: Info = 0.958900714
Information gain: gain = 0.040315165
Revenue:

Table 11: Basic data of customer.

| Numbering | Age | Gender | Profession | Marital status | Income | Area |
|-----------|-----|--------|------------|----------------|--------|------|
| 1         | 35  | Female | Doctor     | Yes            | 8000   | 113.7|
| 2         | 29  | Female | Staff      | No             | 4500   | 95.5 |
| 3         | 40  | Male   | Manager    | Yes            | 7500   | 150.5|
| 4         | 27  | Male   | Teacher    | No             | 5000   | 110.5|
| 5         | 30  | Male   | Civil servants | Yes          | 5500   | 99.8 |
| ...       | ... | ...    | ...        | ...            | ...    | ...  |
| 130       | 33  | Female | Lawyer     | Yes            | 8000   | 125.4|

Table 12: Statistics of age.

| Income range     | Statistical data |
|------------------|------------------|
| Under 3000 yuan  | 23               |
| 3000 yuan-6000 yuan | 59              |
| 6000 yuan or more | 12              |

Table 13: Statistics income.

| Purchase area     | Statistical data |
|-------------------|------------------|
| 120.00m²          | 47               |
| 120.00m² above    | 44               |

Table 15: Statistics of marital status.

| Statistical data |
|------------------|
| Married          | 77               |
| Unmarried        | 12               |

Table 16: Statistics of gender.

| Statistical data |
|------------------|
| Male             | 66               |
| Female           | 25               |

Expected information: Info = 0.710086325
Information gain: gain = 0.289129554

It can be seen that income has information gain in attributes, so it is selected as splitting attribute. Node n is marked with age and grows a branch for each attribute value. Then Yuanzi divided them according to this.

Step 3: generate decision tree and extract rules
According to the above data, a decision tree can be generated. The classification rules are extracted from the decision tree.
R1: if income = middle and occupation = professor then, purchase area = big:
5. Conclusion

With the application of mobile Internet technology, 5G, cloud computing and other network technologies, enterprises have an increasing need for big data, especially in the real estate industry. The era of big data brings real estate marketing not only a challenge, but also an opportunity. Real estate enterprises must seize the business opportunity of big data, adjust their marketing mode in time, and promote the successful transformation and upgrading of real estate enterprises. Based on this background, this paper uses data mining technology to design a precise marketing management system for real estate. Its biggest advantage is to realize sales forecast through mining and analyzing customer data, which provides reference for sales personnel to formulate marketing strategies.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by 2021 Philosophy and Social Science Research Project of Hubei Education Department (PX-321937) and 2021 Project of Hubei Association of Higher Education (PX-321937).

References

[1] S. Li, “Problems and innovative solutions in current real estate project promotion and marketing,” International Journal of Science, vol. 7, no. 2, 2020.
[2] J. Chen, “the era of “Internet plus”, the analysis of real estate marketing countermeasure,” International Journal of Social Science and Education Research, vol. 3, no. 2, 2020.
[3] Y. Yang, “Application of Big Data in Real Estate Marketing,” in Proceedings of the 2019 7th International Education, Economics, Social Science, Arts, Sports and Management Engineering Conference (IEESASM2019), pp. 2527–2530, Dalian, China, August 2019.
[4] W. Bi, J. Wu, Y. Gao, and R. Lin, “Application of creative experiential marketing in vanke real estate,” Journal of Economics and Public Finance, vol. 4, no. 4, 2018.
[5] R. Hou, “Research on the application of consumer psychology theory in real estate marketing,” in Proceedings of the international conference on Economics, Business, Management and corporate social responsibility (EBMCSR 2018) Advances in economics, pp. 142–147, Zhuhai, November 2018.
[6] E. Hromada, “Mapping of Real Estate Prices Using Data Mining Techniques,” Procedia Engineering, vol. 123, 2015.
[7] J. Dong, G. Li, and L. Du, “Research on population distribution model based on real estate big data,” in Proceedings of the 8th International Conference on Management and Computer Science, Shenyang, China, June 2018.
[8] M. Zhuang, W. Pan, Z. Shi, Y. Zhou, and Z. Zhong, “Application of Data Mining Technology in Evaluating Real Estate Investment Plan Based on GRA-AHP,” in Proceedings of the 2019 3rd International Conference on Data Mining, Communications and Information Technology (DMCIT 2019), pp. 295–300, Beijing, China, May 2019.
[9] M. Berry and G. Linoff, “Mastering data mining: the art and science of customer relationship management,” Industrial Management & Data Systems, vol. 100, no. 5, pp. 245-246, 2000.
[10] S. Abdallah, “An intelligent system for identifying influential words in real-estate classifieds,” Journal of Intelligent Systems, vol. 27, no. 2, 2018.
[11] M. Yunita and L. Galinium, “Integrating data mining technique and AHP in market analysis to propose new product development in real estate,” IOP Conference Series: Materials Science and Engineering, vol. 166, no. 1, 2017.
[12] Z. Sen, “The pricing analysis of real estate big data,” International Journal of Intelligent Information and Management Science, vol. 5, no. 4, pp. 70–72, 2016.
[13] E. Hromada, “Real Estate Valuation Using Data Mining Software,” Procedia Engineering, vol. 164, 2016.
[14] N. Bhargava, R. Bhargava, R. P. Singh, and R. Agrawal, Artificial Intelligence and Data Mining Approaches in Security Frameworks, John Wiley & Sons, Hoboken, New Jersey, 2021.
[15] S. Zhang, J. Chen, W. Zhang, Q. Xu, and J. Shi, “Education data mining application for predicting students’ achievements of Portuguese using ensemble model,” Proceedings of the Fifth International Conference on Education and Innovation, vol. 9, pp. 30–34, 2021.
[16] M. Ranjbarfard and S. Ahmadi, “A study of data requirements for data mining applications in banking,” Journal of Digital Information Management, vol. 18, no. 3, 2020.
[17] H. Niavand, “Data mining, applications tools in insurance strategies,” Asian Journal of Management, vol. 11, no. 1, 2020.
[18] A. W. Desta and J. S. Nixon, “Data mining application in predicting bank loan defaulters,” International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 4, 2020.
[19] N. Gulsoy and S. Kulluk, “A data mining application in credit scoring processes of small and medium enterprises commercial corporate customers,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, no. 3, 2019.
[20] I. Sangaiya and A. V. A. Kumar, “A hybrid feature selection method for effective data classification in data mining applications,” *International Journal of Grid and High Performance Computing*, vol. 11, no. 1, 2019.

[21] X. Zhang, S. Wang, C. Gao, and A. Cuzzocrea, “Social Big Data: Mining, Applications, and Beyond,” *Complexity*, vol. 2019, Article ID 2059075, 2 pages, 2019.

[22] S. C. Robinson, "The good, the bad, and the ugly: applying rawlsian ethics in data mining marketing,” *Journal of Media Ethics*, vol. 30, no. 1, 2015.

[23] Y. Ren, S. Wang, Y. Luo, and S. Chen, “ID3 algorithm-based research on college students’ mobile game preferences and analysis of circumvention paths,” *Academic Journal of Engineering and Technology Science*, vol. 3, no. 2, 2020.

[24] G. Qiang, L. Sun, and Q. Huang, “ID3 algorithm and its improved algorithm in agricultural planting decision,” in *Proceedings of the 2nd International Conference on Environmental Prevention and Pollution Control Technologies*, pp. 382–388, Sanya, China, 2020.

[25] E. Edward and O. P. A. Nlerum, “Iterative dichotomizer 3 (ID3) decision tree: a machine learning algorithm for data classification and predictive analysis,” *International Journal of Advanced Engineering Research and Science*, vol. 7, no. 4, 2020.

[26] J. Huo, Q. Kang, and Z. Chen, “Research on the best decision-making recommendation model for airport taxi,” *International Journal of Computational and Engineering*, vol. 4, no. 4, 2019.

[27] I. D. Mienye, Y. Sun, and Z. Wang, “Prediction performance of improved decision tree-based algorithms: a review,” *Procedia Manufacturing*, vol. 35, no. C, 2019.

[28] H. Wang, L. Hu, X. Xie, L. Zhou, and H. Li, “Incremental integration algorithm based on incremental RLID3,” *International Journal of Performability Engineering*, vol. 15, no. 1, 2019.

[29] G. Ma, L. Zhang, G. Cui, and Y. Cheng, “Design of medical examination data mining system based on decision tree model,” *Journal of Physics: Conference Series*, vol. 1237, no. 2, Article ID 022022, 2019.

[30] H. Wang, T. Wang, Y. Zhou, L. Zhou, and H. Li, “Information classification algorithm based on decision tree optimization,” *Cluster Computing*, vol. 22, no. 3, pp. 7559–7568, 2019.

[31] B. V. Sudhakavya, V. Swathi, and S. Senthil, “Classification algorithm in data mining,” *International Journal of Advanced Networking and Applications*, vol. 10, 2019.