The Role of Big Data-based Precision Marketing in Firm Performance

Abstract

Traditional marketing is faced with difficulty in adapting to the changes. In the new business environment, big data is the cornerstone of future business and precision marketing is the main development direction of future marketing. Mining useful data from large amounts of information and applying it to a company's precision marketing will boost the competitiveness and development of the company. This paper is intended to analyze the influence of precision marketing on the corporations' operating revenue conditions, and it takes Company A Digital Marketing as an example. To achieve this purpose, this study develops definite hypotheses based on the cases and theories of precision marketing and shows precision marketing process including data collection, label analysis, page modification and test results. This study uses Decision Tree Regression Analysis using the Python 3.7 and K-means Cluster Analysis by SPSS 24.0 respectively to test the assumptions. Results are revealed and discussed.

Keywords: Precision marketing; Big Data; Digital Marketing; Customer Portrait; Decision Tree Regression Analysis; K-means Cluster Analysis

1. Introduction

In recent years, the continual upgrade of algorithms has promoted the analysis and application of big data, which has caused subversive changes in humanity's previous business models. Big data has brought certain opportunities as well as challenges to modern enterprises and become the next arena of innovation, competition and productivity (Manyika et al., 2011). According to the Capgemini and EMC joint study of approximately 1,000 corporate executives and decision makers from 10 countries in 2015, 65% of respondents believed that companies will risk losing their competitiveness if they do not take proactive measures against big data. 70% of IT decision makers took the view that whether a company can extract value from big data will be closely related to the company's future success. 59% of respondents claimed that data held by companies is becoming an essential part of their market value (Capgemini, 2015).

In the future, the market's trend will demand unique and individualized marketing and enterprises need to capture the right customers and do the right things for them on the right time (Philip, 2007). In order to realize this marketing vision, precision marketing was brought into being. Based on the positioning of precision marketing, enterprises can create digital personalized service advertisements, achieving high efficiency and low cost goals (Xu, 2006). With the advancement of big data technology, data brings more interaction, integration, exchange and transaction. Companies can incorporate big data into their marketing plans, making marketing targeted, accurate and efficient and improving their core competitive edge (Heini and Heikki, 2015).

This paper looks at how precision marketing strategies based on big data influences sales revenue and what other companies can learn from Company A's precision marketing strategies. To answer these questions, this study processes existing user data through Decision Tree Regression Analysis Model and K-means Cluster Analysis Model to establish a customer tag database and user portraits, and implement precise marketing by optimizing pages. Finally, conclusions are drawn from the comparison of revenue data before and after implementing precise marketing strategies.

2. Literature Review

Precision marketing theory proposes that: enterprises need to a) adopt more precise and measurable marketing methods, b) orient marketing plans towards results and action, c) and focus on direct marketing investment (Philippe, 2011). Consumers' costs can be reduced through precise marketing, because precision marketers provide a wealth of information of goods and services for customers, reducing their trouble and time cost of shopping. Precision marketing also helps customers quickly compare prices and information of goods or services to decide whether they purchase them or not. In this way, the theory let marketers see ways of decreasing marketing costs and improving the marketing efficiency.

2.1 Impact of Big Data on Precision Marketing

Precision marketing involves data collection, data processing, and analysis of user data. It can be stated that data plays a vital role in precision marketing. At the same time, precision marketing draws on research methods in social psychology, behavioral science and modern business management to comprehensively explain the user’s inner need. The organic combination of subjective and objective methods allows marketing to correctly identify market needs.

In the era of big data, data information becomes highly important. Until the user purchases and eventually becomes a member, data collection and mining analysis are needed for decision-making in the entire marketing process. Generally speaking, the basic marketing idea in the context of big data is to collect and analyze user data, including gender, occupation, age, marital status, income, user browsing behaviors, purchase behavior, etc. Through the collection and analysis, basic portraits of users are established to grasp their needs, according to which differentiated products and precise marketing strategies are developed. Finally, these trigger consumer purchase (Ronald, 2012). Wen and Zhu (2015) studied the relationship between big data and precision marketing in the book "IT to DT: Big Data and Precision Marketing", and pointed out that companies employing big data mining technology can achieve targeted marketing strategies. Li Jun (2015) argues that companies use big data to accurately identify customers in marketing, which prompts companies to achieve precision marketing and improve the overall efficiency of companies. In the view of above scholars, grinding big data technology is the cornerstone of precision marketing. Big data can provide insights into market behavior and guide enterprise marketing activities, providing a technical pivot for the transformation of marketing.

2.2 Impact of Precision Marketing on firm performance

Precision marketing is a marketing activity using quantitative analysis of the market, modern information technology and individualized communication technology, such as database, customer relationship management (CRM), modern logistics, etc., to maximize enterprises' benefits (Jeff and Gresh, 2008). Precision marketing realizes long-term customized communication with consumers in virtue of modern information technology and network communication technology, turning the enterprise-consumer communication to a continuous and effective process (Gai, 2015).

The merits of precision marketing also include reducing enterprises' high costs in traditional advertising, making their marketing activities direct, measurable, controllable and effective, really meeting customers' need, and promoting the rapid development of companies. It creates a contact zone where enterprises and customers can directly interact with each other. Therefore, it can be stated that precision marketing is the future success of the company. Several cases prove its positive impact on domestic and foreign companies.

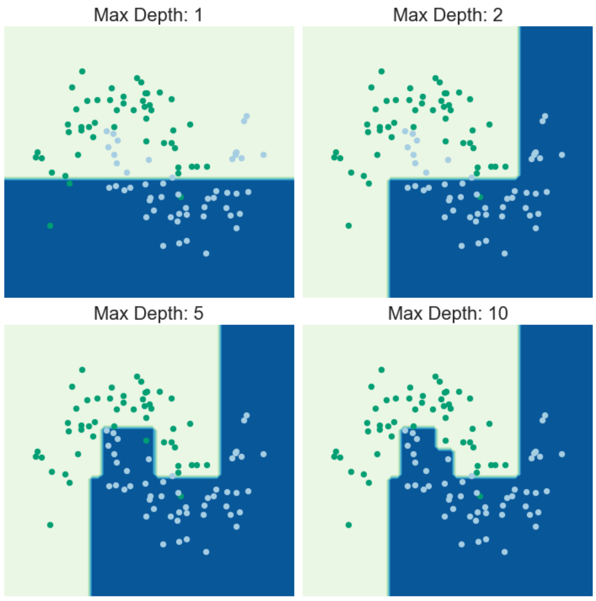
Based on the research discussed above, this study proposes the following hypothesis: H1: Precision marketing has positive impacts on company sales revenue.

2.3 Using Business Intelligence to Achieve Precision Marketing

2.3.1 Influence Factors Based on Decision Tree Regression Analysis Model

Decision Tree, a method of analyzing decisions, evaluates project risks and determines the probability that the expected value of NPV exceeds or equals zero (Miomir et al., 2017). In this model, the factors that are most likely to influence outcome are the most important. Using this model, companies can quickly identify the factors affecting the marketing results in the existing data, so as to prepare and lay the foundation for the subsequent customer portraits.

Figure 2-1: Decision Tree Regression Analysis Model



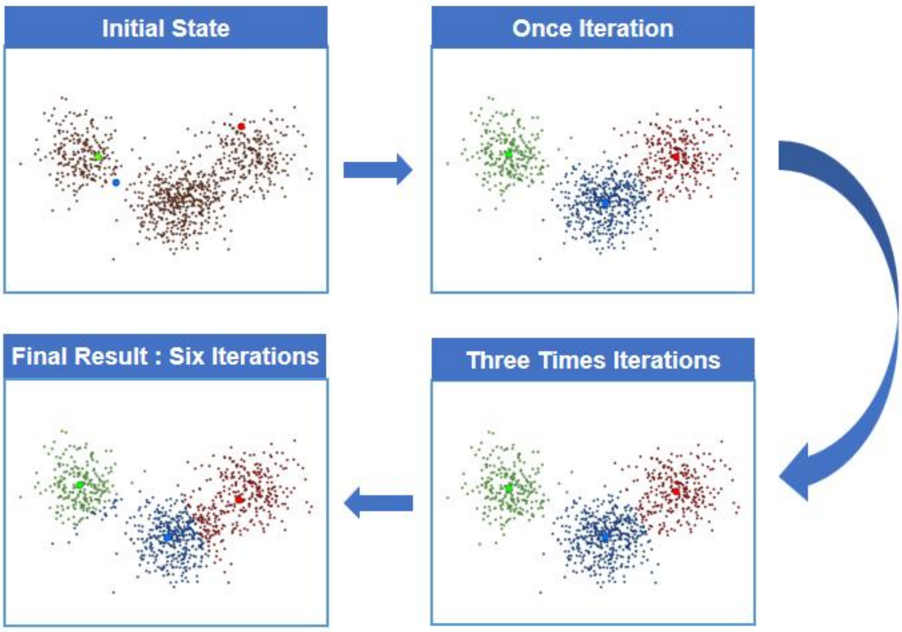
Source: Pivotal Engineering Journal, Interpreting Decision Trees and Random Forests (Greg, 2017)

2.3.2 Customer Value Hierarchical Based on K-means Cluster Analysis Model

In 1967, the K-means algorithm, one of the most commonly used cluster analysis algorithms, was introduced by MacQueen. Its principle is to set the number of categories “k” as input parameters, divide the data into “k” clusters, and determine which cluster a data object should be merged into according to the distance between the data object and the cluster centers. Data pairs in the same cluster objects are very similar, and the data object dissimilarity between different clusters is more noticeable (MacQueen, 1967).

Cluster analysis classifies a collection of physical or abstract objects into multiple groups of similar objects. It is a statistical analysis method for studying classification problems, and it is also an important data mining algorithm (Xue, 2018). Cluster analysis is comprised of a range of patterns. A pattern is a vector of measurement or a point in a multidimensional space. Cluster analysis is built on similarity, and patterns in one cluster have more similarities than those in different clusters. Algorithms for cluster analysis can be divided into Partitioning Methods, Hierarchical Methods, Density-based Methods, Grid-based Methods, and Model-Based Methods.

Figure 2-2: K-means Cluster Analysis Model



2.3.3 User Portrait Based on Analysis Models

A user portrait refers to a model drawing on a variety of types of data information to comprehensively describe a specific user. Each piece of specific information of a user needs to be abstracted into a label. The tags together describe a user's image concretely and help tailor services to the user. The process of building up user portraits includes collecting source data, establishing customer label database, creating user models and predicting user behaviors (Fei, et al., 2018). Big data is used in the process, together with multiple other technologies including collection, cleaning, processing and mining. Establishing user portraits is the foundation of precision marketing. Effective user portraits can make products focused on customers, improve the effectiveness of decisions, and thus promote marketing efficiency.

3. Methodology

3.1 Research Description

This research deduces the precise marketing process of Company A. The user behavior data and basic information of users are analyzed by means of Decision Trees Regression Analysis Model and K-means Cluster Analysis Model to work out precise marketing strategies and examines the enterprise earning changes to investigate whether precise marketing has a positive effect on corporate revenue.

3.2 Research Design

3.2.1 Data Collection

The data used in this study are extracted from the company's internal database. The research object is the users who had web behaviors on the official website of Company A in the second and third quarters of 2019. By linking the user ID, process tracked URLs (eg: http://www.aaaaa.com/security/?cm_mmc=Email+Security_Network+Protection/Protect+your+network+layout2&cm_mmca1=000000Ml&cm_mmca2=10000099&cm_mmca3=M00000319) which are extracted from the company's internal database to generate basic attribute data and user behavior data (all the data have been desensitized).

The whole data are collected from two sources. One is the consumer database including customer consumption records, web behaviors, product preferences, jobs and industries. The other is the company's marketing database including product revenue, regional revenue, sales volume and sales volume growth. The original data is collated, because the sheer scale of the data extracted from the system is quite large, and some segments do not comply with the purpose of this study, .

Finally, this research has obtained the purchase records of the research object in each quarter since 2016, a total of 13,621, divided into 156 dimensions, including Behavior code (BLD_KEY), Date (RPTG_DATE), Business opportunity profit (PIPE_REV_AMT_MIL), etc.

3.2.2 Finding Influence Factors Based on Decision Tree Regression Analysis Model by Python 37.0

Examining the natural attributes of the customers is most likely to affect the transaction through the decision tree and help marketers filter out valueless customer information and increase productivity. The research question in this analysis is: What factors about users' information influence revenue? To answer the question, this article imported thirty nine discrete variables into the model based on survey results.

Proposed by Quinlan in 1986, decision Tree ID3 is a well-known algorithm in machine learning and also the earliest and most influential decision tree method proposed in the world (Yin and Sun, 2018). The basic principle is identifying the appropriate attribute in the entire data and using it as the root node of the decision tree model, and then finding other attributes along the decision tree. The other attributes in the data are sequentially reviewed and classified. In the Decision Tree ID3 algorithm, the choice of attributes is the most important. The attribute with the maximum gain is the root node of the decision tree and needs to be defined by entropy, one of the concepts in information theory (Zhang, 2013). The information gain of entropy is a criterion for determining the information attribute. The selection of information entropy is the basis of optimal test attribute of ID3 algorithm.

The information entropy formula is as follows, in which 'p_k' stands for the proportion of the sample k in the data of the current node D.

$$Ent(D) = -\sum_{k=1}^{|\gamma|} p_k \log(p_k) \tag{1}$$

The entropy value is inversely related to the property change.

The information gain formula is:

$$Gain\,(D,a) = Ent(D) - \sum_{V-1}^{V} \frac{|D^V|}{|D|} Ent(D^v) \quad (2)$$

The information gain value and the purity of the attribute are inversely proportional (CSDN, 2014).

ID3 algorithm has a clear theory, a simple structure, a strong learning ability and fast classification. Especially, it is well applicable to large-scale data classification. According to the principle of decision tree algorithm, results can show the main factors affecting the sales revenue.

3.2.3 Achieving Customer Value Hierarchical Based on K-means Cluster Analysis Model by SPSS 24.0

The purpose of this model is to analyze the more important factors which are worked out by the decision tree. With the help of the SPSS 24.0 statistical tool, the data obtained are clustered in several groups. According to the result data, customers are divide into three value segments, high, medium and low, by evaluating their value. Finally, the final data is submitted to the sales area in the form of a detailed list to provide a basis for the market personnel in the sales area.

Next, the "final data" mentioned in the previous paragraph will be divided into k clusters, which is preset as input parameters, and all the data objects are put into different clusters according to their Euclidean distances to the cluster centers Go. The data objects in the same cluster are very similar to each other, and the data objects between different clusters are distinct. After the clustering is completed, the data is matched to the corresponding customers' IDs, and the customers are labeled as high, medium or low value.

3.2.4 Optimize Marketing Strategies Based on Analysis Model results

Employees of Company A shifted their focus to customers with high-value labels (K-means), collected high-value customers' preference data labels, and conducted precise marketing on them. Two channels are employed. First, the IP of high-value customers are marked, and accurate push notifications of pre-set advertisements and product introductions are sent. When the user logs in to the company's official website for the second time, the backend system will provide the user with a customized page based on the user’s IP address. Second, the customer service staff conduct secondary sales to high-value customers through telephone and email.

3.2.5 Forest Conclusion

If a clear increase in revenue data is observed when the revenue data before and after the implementation of precision marketing are compared, it proves that the hypothesis is true, and vice versa.

4. Precision Marketing Process

4.1 Precision Marketing Frame

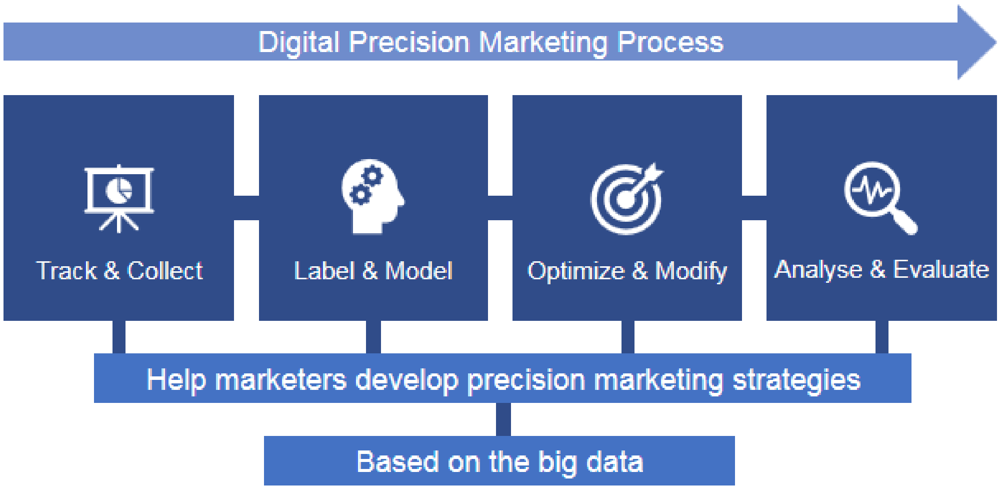
“Track & Collect”: P platform of Company A is used to collect the real-time and non-real-time data of customers, and the data of third parties are integrated through the link platform. In this process, P platform turns the raw data collected into Dashboard Mapping.

“Label & Model”: Machine learning and Desicion Tree Models are used to help Company A gain deep insights in customer data. Company A organizes customer information, builds a customer tag library, maps customers with tags, and finally establishes user portraits.

“Optimize & Modify”: P intelligent data platform uses user portraits as the basis for judgment users’ value, optimizes and improves webpages of the company through Drupal, so as to push customized advertisements to customers. At the same time, marketers make improvements to market strategies

“Analyse & Evaluate”: By comparing the differences in feedback data from the platform, Company A can judge the impact of precision marketing on current marketing revenue.

Figure 4-1 Process of Precision Marketing



4.2 Data Requirement and Collection (tracking & collecting)

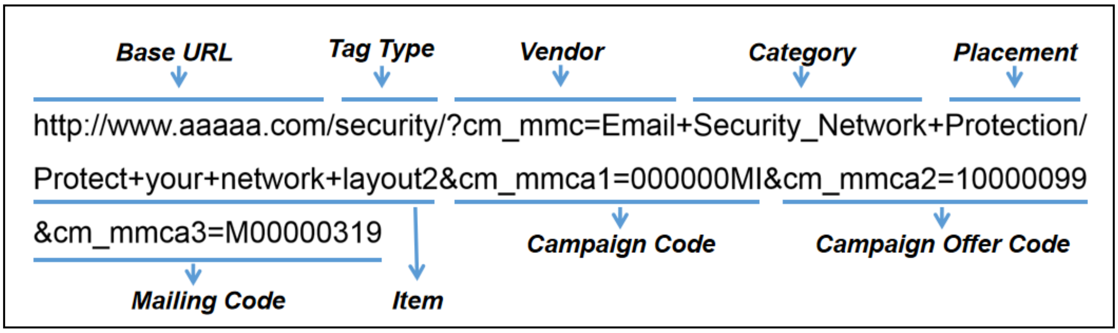
4.2.1 Data Requirement in Precision Marketing

A huge amount of raw data is the basic condition for constructing a precise marketing data platform for management products which need to be collected and stored. Raw data mainly includes personal information of customers, geographical location information, URLs on the upper and lower levels of web pages, and product information.

4.2.2 URL Tracking (user's access behavior path in the website)

Several questions need answering for a given page (URL), including where a user to this page comes from, where the user goes after entering this page, and whether the user's navigation path follows or deviates from the path designed by the prediction of backend systems. User behavior path analysis, an analytical method, can offer answers to them. In additional, it guides the operation to clarify the existing user path, optimize user behaviors along the optimal access path, and adjust the front-end layout according to the needs in business scenarios. In this process, data are collected, including personal information such as IP address.

Figure 4-2 Example of Tag



An analytic tool called Coremetrics is used in the following step. Coremetrics inspects companies' capabilities of tracking user activities on web pages and digital assets and tagging parameters including VCPI (Virtual Control Program Interface) to provide detailed information on where a person comes from. Thus, user activity data are made available to P platform for enabling nurturing campaigns.

Figure 4-3 User Path (Original-Targeting)

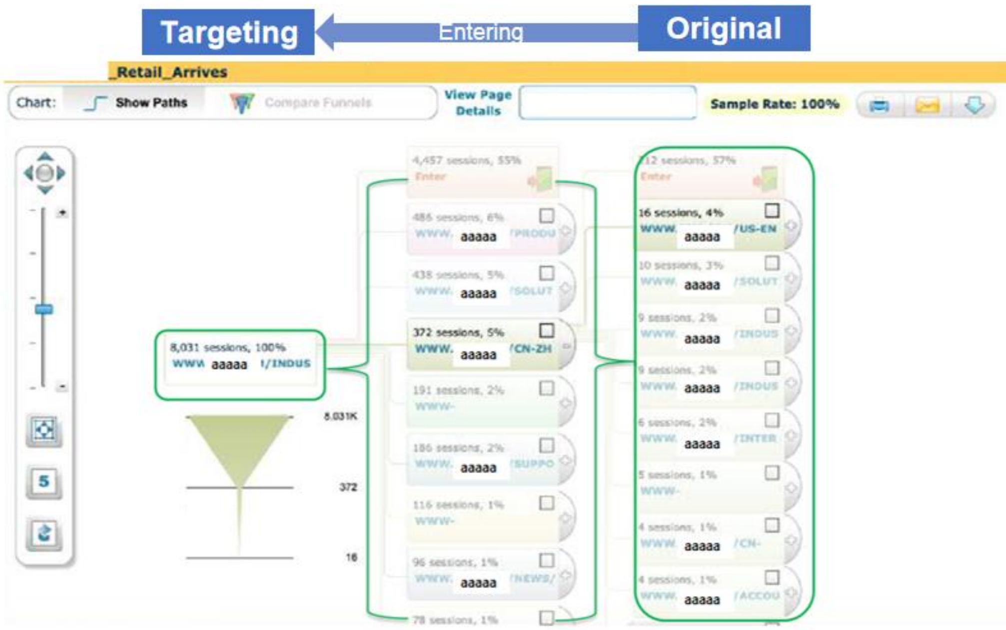
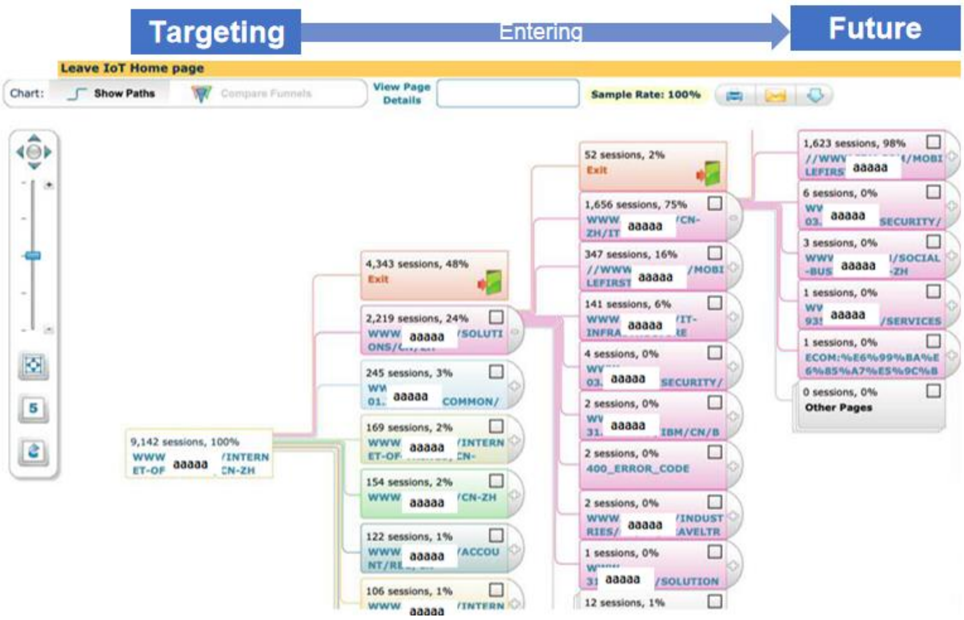


Figure 4-4 User Path (Targeting-Future)



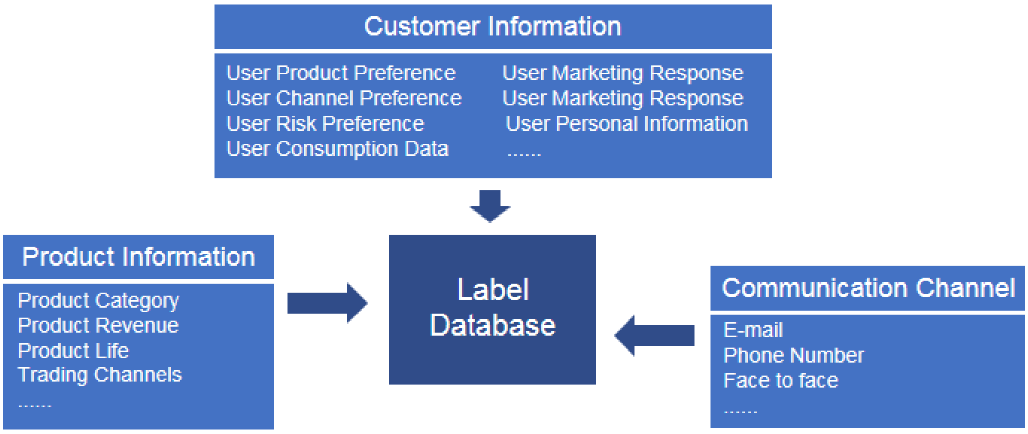
4.3 Building Customer Tag and User Portraits (labeling & modeling)

According to information collected, what products should be recommended to users can be found out. To fulfill this aim, data in bulk are analyzed, including users' basic information, online browsing information, purchasing information, behavior trajectory, lifestyle, and habits. In doing so, users' behavior characteristics are described and their portraits are continuously improved. In one word, tracking users' complete life cycles refines precise marketing strategies of products.

4.3.1: Establishing Customer Label Database and User Portraits

A customer label database includes customer data and product data. Based on the big data technology, a label database integrates and analyzes raw data to identify each customer's characteristics. Then, a comprehensive user portrait is to be formed by combining different customer labels. The basic principle for establishing a customer label database is comprehensiveness and authenticity of customer labels. Comprehensiveness means that building up a label needs to cover all characteristics of a customer and pay attention to all the data of the customer's entire life cycle. Authenticity refers to ensuring that every label in a label database is authentic and valid; otherwise, incorrect label information will affect the reliability of user portraits. A comprehensive and authentic label database can improve the efficiency of creating user portraits and ensure the success rate of precision marketing.

Figure 4-5 Label Database Frame



After a customer label database is established, customers are mapped to labels to form user portraits. The key to forming a user's portrait is combining multiple labels. Broad label coverage leads to an accurate description of a user. The types of labels covered in this article are as follows:

a) Statistical Labels: This type of labels are the most basic, consisting of users' gender, jobs, geographic locations, etc. The information can be derived directly from user registration data, user access and consumption data.

b) Rule Labels: This type of labels are generated based on user behavior and established rules. "Assist to Create" is an illustration. After a user communicates for the first time with customer service online through digital advertising, if a sale is made within 90 days, it will be counted as an "Assist to Create", otherwise it will be regarded as other sales types. When developing portraits, digital marketers are familiar with the business and familiar with the structure, distribution, and characteristics of the data. Therefore, the rules of the rule-type label are determined by digital marketers.

Machine Learning Mining Labels: This type of labels are generated by data mining and used for predicting and judging certain attributes or behavior of the users, for example, judging a user's preference for a certain product according to the user's consumption habits. This type of labels need to be generated by algorithmic mining a step-by-step basis.

4.3.2: Decision Tree Regression Analysis Model

The Decision Tree ID3 algorithm is a classification prediction algorithm introduced by J. Ross Quinlan. Its basic steps are: the choice of test attributes is determined by information gain value. The segmentation of a sample set is determined by the test attributes chosen. Different values segment the sample set into several sub-sample sets. In the meantime, the node corresponding to the sample set of the decision tree generates a new child node.
The information entropy formula and information gain formula are:

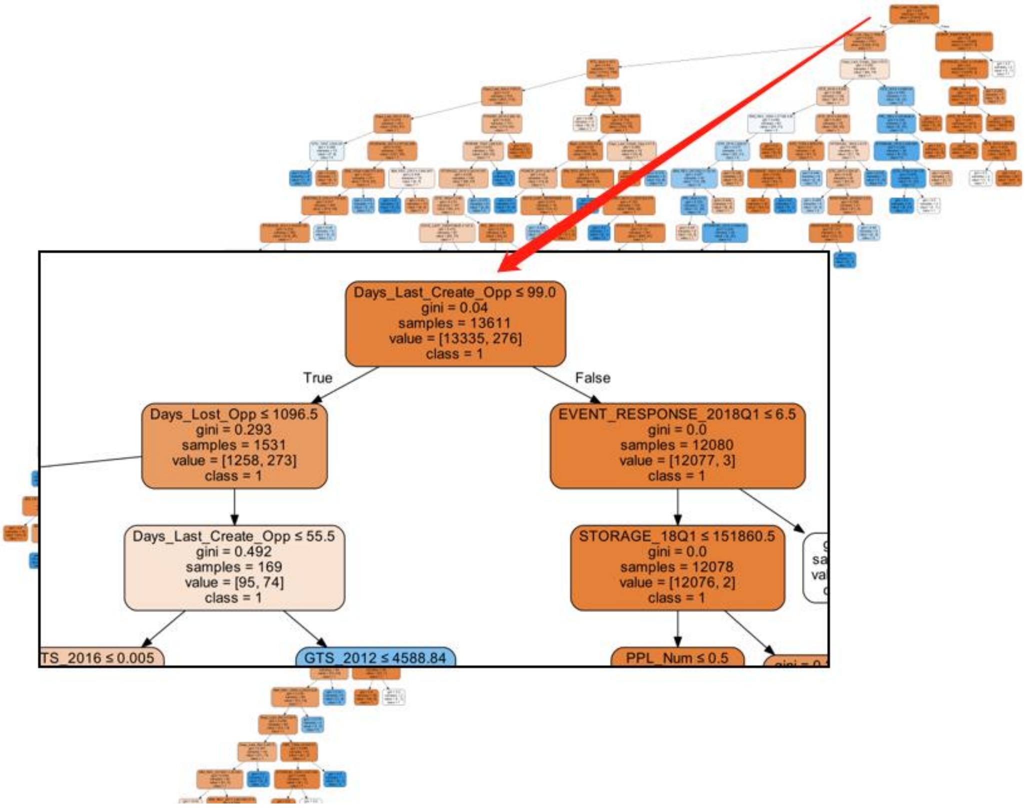
$$Ent(D) = -\sum_{k=1}^{|\gamma|} p_k \log(p_k) \tag{1}$$

$$Gain(D,a) = Ent(D) - \sum_{V-1}^{V} \frac{|D^V|}{|D|} Ent(D^v) \tag{2}$$

The entropy value is inversely proportional to the property change. This relationship applies to that between the information gain value and the purity of the attribute (CSDN, 2014).

To find the natural attributes of customers which are most likely to affect transactions, this paper uses the decision tree to analyse 13,621 data which are divided into 156 dimensions. The results obtained by Python show that ‘Create_ within 99 days’ is the most influential factor among the natural attributes. 97.9% of the order records have the common natural attributes of ‘Create_ within 99 days’. The influence of other factors decreases layer by layer, and the underlying factors are not considered in this project.

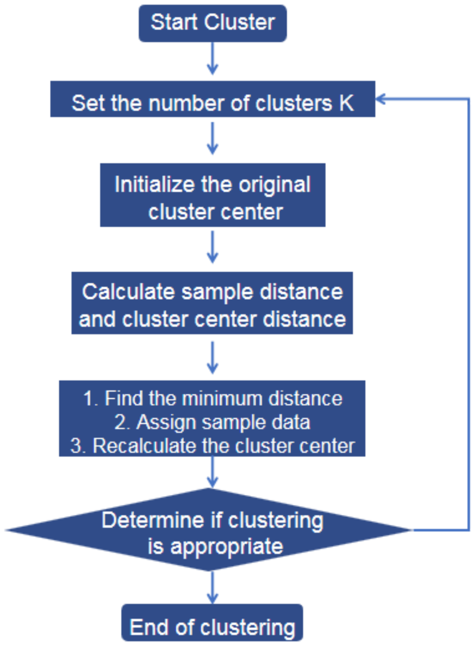
Figure 4-6 Decision Tree Regression Analysis Model Result



4.3.3 K-means Cluster Analysis Model

The k-means algorithm was proposed by MacQueen in 1967, as one of the most commonly used cluster analysis algorithms (MacQueen,1976). Firstly, existing data is processed into a scatter plot. K objects are randomly selected in a database as the initial center points of K clusters, and then the distances from the rest points to the K center points are traversed. An object is put into one group closest to one of the center points, and then the mean of all objects in each group is calculated to be the new center of the group. If a group center no longer changes or reaches the number of iterations, the algorithm ends, otherwise it returns to the step of distance calculation. The flowchart of the algorithm is as follows.

Figure 4-7 K-means Algorithm



According to the results of the decision tree, it can be seen that the “time” factor plays an important role in marketing, which includes the last time to browse web products, the last time to contact a telemarketer, and the last time to buy products. In K-means model, the K-means model can be used to perform a cluster analysis on the relevant factors mentioned above in SPSS 24.0, so as to divide the customer value into three value levels of high, middle and low. This study selects the target customers’ " the last time to browse web product from the present" (DAYS_LAST_RESPONSE), "the last time to contact telemarketers from the present" (Days_Last_Create_Opp), "the last time to buy the product from the present" (Days_Last_Win). These index parameters are used as dimensions in the cluster analysis. The data used are clustered after being cleaned up. According to the results of the decision tree, this research selects the “Days_Last_Create_Opp” as an optimal property.

Table 4-1

Customers Clustering Index	
the last time to browse web product from the present	DAYS_LAST_RESPONSE
the last time to contact telemarketers from the present	Days_Last_Create_Opp
the last time to buy the product from the present	Days_Last_Win

With the help of the SPSS statistical tool, the data are clustered in 15 groups according to the dimensions in the table above, and the high, medium and low value segments of the customers are formed. The results are shown below:

Table 4-2

Final clustering center

	cluster														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
DAYS_LAST_RESPONSE	550	783	10	35	541	22	20	426	14	28	19	685	21	650	464
Days_Last_Create_Opp	42	28	39	38	38	44	38	50	53	43	40	91	42	44	40
Days_Last_Win	625	5000	2905	892	2887	1765	104	137	2361	460	5000	2111	1255	1542	1116

Table 4-3: Number of Cases in Each Cluster

Cluster	1	20.000
	2	2.000
	3	16.000
	4	202.000
	5	4.000
	6	74.000
	7	617.000
	8	15.000
	9	29.000
	10	276.000
	11	118.000
	12	1.000
	13	131.000
	14	6.000
	15	20.000
Effectiveness		1531.000
Missing		.000

The final clustering result of analyzing" the last time to browse web product from the present" (DAYS_LAST_RESPONSE), "the last time to contact telemarketers from the present" (Days_Last_Create_Opp), "the last time to buy the product from the present" (Days_Last_Win) is as follows:

Table 4-4

Value clustering results

Clustering Center	Number of Cases	Proportion	Value type
1	20	1.31%	low
2	2	0.13%	low
3	16	1.05%	low
4	202	13.19%	middle
5	4	0.26%	low
6	74	4.83%	low
7	617	40.30%	high
8	15	0.98%	low
9	29	1.89%	low
10	276	18.03%	middle
11	118	7.71%	low
12	1	0.07%	low
13	131	8.56%	low
14	6	0.39%	low
15	20	1.31%	low

From the results of user stratification, it can be seen that the common feature of users is that the time interval from browsing products to purchase products is short and concentrated within 200 days from the last purchase. Therefore, when future customers show similar network behavior characteristics, salespeople should pay attention to them.

4.4 Optimizing web advertising settings (optimizing & modifying)

Web optimization: it consists of all the technical and marketing techniques of learning traffic and engaging visitors, and then converts them into customers and prospects. It uses a clearly defined method meeting all pre-defined goals. There are several optimization types:

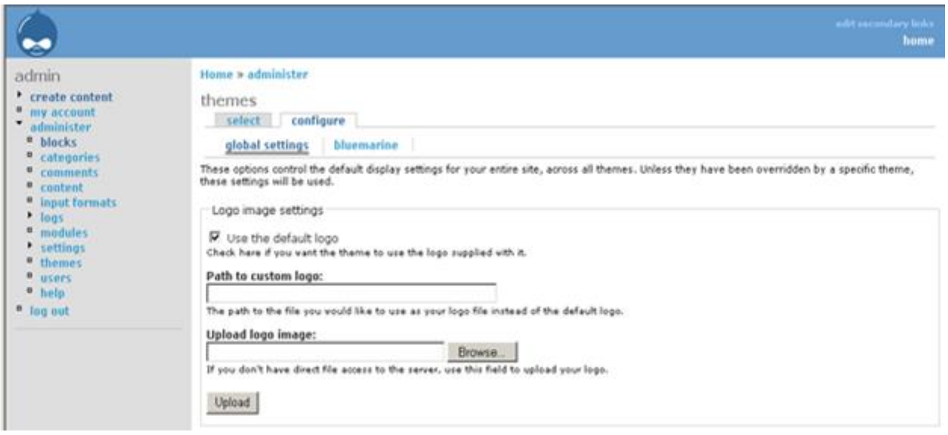
a) Search engine optimization: it includes website content and link modification as well as on-page optimization.

b) Technical platform optimization: it focuses on site speed, display on mobiles and tablets, page errors and hosting performance.

c) Conversion & user experience optimization: it involves User Session Recording, Heatmaps, surveys and user feedback, Visitor Segmentation, and Advanced A/B testing.

After completing the customer tagging work, marketers customize web pages for customers with high-value tags based on the results provided by the models above. Drupal is used to optimize the layout of the pages and the push of advertisements according to the rules of user behavior preferences. Used for building websites, Drupal is a highly modular, open source web content management framework that focuses on collaboration. It is a system that is extensible, adapts to the standards, and strives to maintain concise code and smaller scripts. A Drupal release includes basic core functions, and other additional functions can be obtained by installing modules. In addition, marketers need to make corresponding marketing policy adjustments in telemarketing and email marketing.

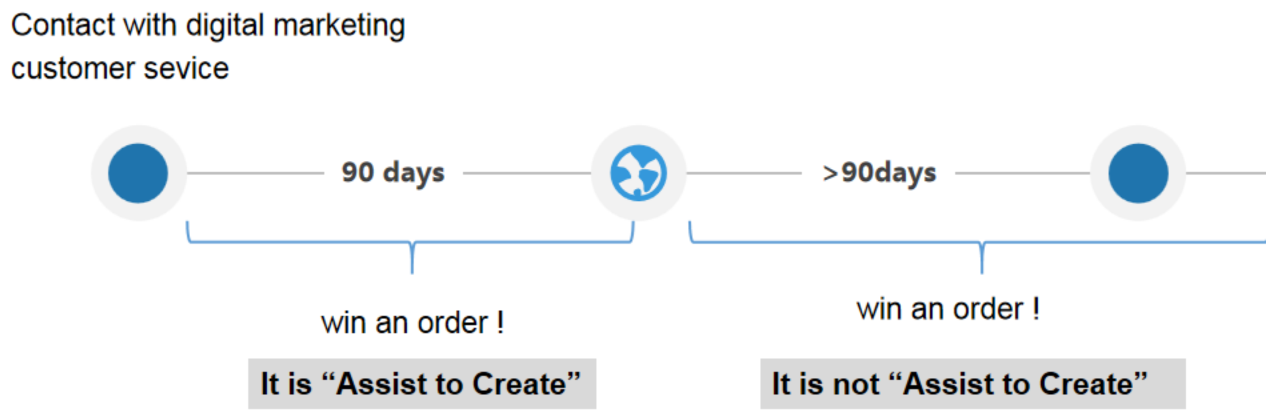
Figure 4-8: Number of Cases in Each Cluster



4.5 Analyzing result data

Monitoring data: it means monitoring effective “Assist to Create” information and profit comparison before and after the implementation of precision marketing strategies. The detailed analysis process will be explained in the next chapter.

Figure 4-9: Assist to Create



5. Results and discussion

5.1 Result of ‘Assist to Create’ marketing source data

Explanation of the words in the table:

a) VLC is the number of valid users

b) VLRC is the revenue of valid users (prediction). Potential revenue associated with a validated lead is measured as validated lead revenue created (VLRC).

c) Win is the revenue of orders. When a deal is closed, it is called a win. Wins happen at a lead level. An opportunity can have multiple wins as well as losses, so this part of the funnel is not necessarily smaller than the opportunity step.

d) Win Count is the number of orders

e) Win Revenue is the revenue associated with wins. This number can fluctuate, and a win in one reporting period may ultimately become a loss.

The two tables below show the performance of Company A's precision marketing strategy for the ‘Infrastructure Service’ product category in mainland China, Hong Kong, Macao and Taiwan in the second and third quarters of 2019, including the value of effective customer information and contribution value of company sales revenue.

Figure 5-1: 2019 Q2 'Assist to Create' Marketing Source Data

Business UnitGTS **Geo Level 1**Greater China Group **Mktg Influence Type**Assist to Create

Marketing Sunrise Reporting Revenue Value is Capped Data as of 2019 Q2		Marketing Source			
		VLC	VLRC $M	Win Count	Win
UT Brand Level 1	Geo Level 3	Actual	Actual	Actual	Actual
Infrastructure Services	China	22	$10.70	5	$3.60
	Hong Kong	8	$6.50	5	$0.70
	Macao	–	–		–
	Taiwan	12	$2.30	3	$0.40
	Summary	42	$19.50	13	$4.70
Technology Support Servi	China	64	$10.10	12	$1.60
	Hong Kong	15	$1.10	14	$0.90
	Macao	6	$0.20	–	–
	Taiwan	16	$2.20	3	$3.30
	Summary	101	$13.60	29	$5.80
Summary		143	$33.10	42	$10.40

Figure 5-2: 2019 Q3 'Assist to Create' Marketing Source Data

Marketing Sunrise Reporting Revenue Value is Capped Data as of 2019 Q3		Marketing Source			
		VLC	VLRC $M	Win Count	Win
UT Brand Level 1	Geo Level 3	Actual	Actual	Actual	Actual
Infrastructure Services	China	45	$33.80	6	$3.40
	Hong Kong	34	$32.10	2	$0.80
	Macao			–	
	Taiwan	11	$4.40	3	$0.80
	Summary	90	$70.20	11	$5.00
Technology Support Servi	China	113	$38.80	12	$7.40
	Hong Kong	30	$2.20	9	$1.10
	Macao	6	$0.20	–	–
	Taiwan	16	$2.80	12	$0.20
	Summary	165	$44.00	33	$8.70
Summary		255	$114.20	44	$13.70

5.2 Result of "Assist to Create" changes data

The table below shows the impact of digital precision marketing on corporate sales revenue. 'Indicator' means the estimated value of 'Assist to Create' for this quarter based on the situation of the previous quarter. 'Actual' indicates actual precision marketing's contribution to corporate sales revenue. The green color means that after marketing strategies are changed through precision marketing, the contribution of digital platforms to company earnings has increased.

Table 5-1: 2019 Q3 'Assist to Create' Changes Data

		'Assist to Create' Win Revenue			
YTD: 19Q3		**Campaign**	**Indicator**	**Actual**	**Attain%**
	Hybrid Cloud	WCP	4.49	1.17	26%
	Hybrid Cloud	HCI - Integration & Development Management	2.35	7.64	325%
	Hybrid Cloud	HCI - Digital Business Automation	0.48	0.39	82%
China	Hybrid Cloud	HA - ICP-Data	1.52	-	0%
	Hybrid Cloud	HA - Hybrid Data Management	1.52	5.29	349%
	Hybrid Cloud	Narrative	-	0.81	-
	Hybrid Cloud	Developer	-	-	-
	Hybrid Cloud	**Hybrid Cloud Total**	**11.87**	**18.77**	**158%**
	Hybrid Cloud	WCP	0.96	1.62	168%
	Hybrid Cloud	HCI - Integration & Development Management	0.22	1.00	452%
	Hybrid Cloud	HCI - Digital Business Automation	0.05	0.16	353%
Hong Kong	Hybrid Cloud	HA - ICP-Data	0.14	-	0%
	Hybrid Cloud	HA - Hybrid Data Management	0.14	0.23	159%
	Hybrid Cloud	Narrative	-	-	-
	Hybrid Cloud	Developer	-	0.17	-
	Hybrid Cloud	**Hybrid Cloud Total**	**1.66**	**4.75**	**286%**
Taiwan	Hybrid Cloud	WCP	0.96	4.40	457%
	Hybrid Cloud	HCI - Integration & Development Management	0.19	6.12	3163%

	Hybrid Cloud	HCI - Digital Business Automation	0.04	-	0%
	Hybrid Cloud	HA - ICP-Data	0.12	0.07	55%
	Hybrid Cloud	HA - Hybrid Data Management	0.12	4.92	3941%
	Hybrid Cloud	Narrative	-	-	-
	Hybrid Cloud	Developer	-	-	-
	Hybrid Cloud	**Hybrid Cloud Total**	**1.57**	**19.53**	**1244%**
	Hybrid Cloud	WCP	6.42	7.19	112%
	Hybrid Cloud	HCI - Integration & Development Management	2.76	14.75	534%
	Hybrid Cloud	HCI - Digital Business Automation	0.57	0.55	98%
Total	Hybrid Cloud	HA - ICP-Data	1.78	0.07	4%
	Hybrid Cloud	HA - Hybrid Data Management	1.78	10.44	585%
	Hybrid Cloud	Narrative	-	0.81	-
	Hybrid Cloud	Developer	-	0.17	-
	Hybrid Cloud	**Hybrid Cloud Total**	**15.10**	**43.05**	**285%**

5.3 Discussion of data results

By comparing the data before and after the implementation of precision marketing, the following observations can be made:

a) VLC increased from 143 in Q2 to 255 in Q3, achieving a 7.83% increase.

b) VLRC increased from US $ 33 to US $ 114.2 during Q2, achieving an increase of 246%.

c) Win Count increased from 42 to 44 during Q2, achieving a 4.7 percent increase.

d) Win increased from US $ 10.4 to US $ 13.7 in Q2, achieving a 31.7% increase.

The growth of VLC and VLRC indicates that digital precision marketing is of great help for companies to deliver right advertising to right customers. Their growth, however, was much larger than that of Win because the influence of digital precision marketing may not be shown in short term. It is necessary to continue observing changes in data in the future. In one word, the company has achieved rapid user conversion by precision marketing. It is believed that the customer purchasing possibility would increase in marketers' follow-up service to customers. This can be demonstrated in the feedback data changes in the subsequent stage. The results show that precision marketing can promote the profitability of enterprises.

The 'Actual' value of 'Assist to Create' can reflect the degree of precision marketing participation in product sales. The sales of multiple products of the 'Hybrid Cloud' production line in China, Hong Kong and Taiwan are examples. From the table "Assist to Create Win Revenue", the 'Actual' value of 'Assist to Create' for most products was higher than expected. Because the sales of a single product line is small, this paper collected data on 'Assist to Create' for more than 70 products. In conclusion, after web page advertisements were modified through precision marketing strategies, the actual revenue value of 'Assist to Create' in this quarter increased.

Therefore, the hypothesis that precision marketing positively impacts company revenue is true.

6. Summary

This paper carries out an analysis of precision marketing, big data and companies' revenue. The research, analysis and optimization are based on Decision Tree Regression Analysis Model and K-means Cluster Analysis Model. First, it discusses the theoretical basis of precision marketing, big data marketing and the general situation of Company A's precision marketing strategies. Then it analyzes processes, methods and advantages of precision marketing strategies by looking into the collection of customer data, customer labels and customer portraits with Decision Tree Regression Analysis Model and K-means Cluster Analysis Model. Finally, the results are revealed and discussed.

References

Anand R. and Jeffrey D.U. (2011). Mining of Massive Datasets, *Cambridge University Press.*

Ariker, M., Heller, J., Diaz, A. and Perrey, J. (2015). How Marketers Can Personalize at Scale, *Harvard Business Review*.

Capgemini (2015). New global study by Capgemini and EMC shows Big Data driving market disruption, leaving many organizations fearing irrelevance.

CSDN (2014). Learn machine learning algorithm: ID3 Algorithm of Decision Tree [Online]. Available at: https://blog.csdn.net/google19890102/article/details/28611225.

Datayuan (2017). Ctrip real-time user data collection and analysis system [Online]. Available at: http://www.datayuan.cn/article/11942.htm.

Deighton, J. and Kornfeld, L. (2009). Interactivity's unanticipated consequences for marketers and marketing, *Journal of Interactive Marketing*, Vol. 23 No. 1, pp. 4-10.

Fei, P., et al. (2018). A Multi-view Fusion Framework for Constructing User Portraits, *Computer Science,* Vol. 45 No. 1, pp.179-182.

Gai, N. (2015). Research on Brand Building Strategy Based on Precision Marketing

Greg, L., (2017), Interpreting Decision Trees and Random Forests, *Pivotal Engineering Journal.*

Heini, M. T. and Heikki, K. (2015). The usage of digital marketing channels in SMEs, *Journal of Small Business and Enterprise Development*, Vol. 22 Issue: 4, pp.633-651.

Jeff Z. and Gresh B. (2008). Precision marketing：the new rules for attracting, retaining, and leveraging profitable customers.

Li, J. (2015), Practical big data: customer positioning and precision marketing, *Tsinghua University Press*

Liu, Y., Mezei, J., Kostakos, V. and Li, H. (2017). Applying configuration analysis to IS behavioral research: a methodological alternative for modelling combination complexities, *Information Systems Journal,* Vol. 27 No. 1, pp. 59-89.

Maex, D. and Brown, P.B. (2013). Big Data Marketing: Positioning Customers, *Beijing Mechanical Engineering Press*, pp. 216-217.

MacQueen, J.B. (1967). Some Methods for classification and Analysis of Multivariate

Observations: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. *University of California Press.* pp. 281–297.

Manyika J., et al. (2011). The next frontier for innovation, competition, and productivity, New York:Mc Kinscy Global Institute.

McCarthy, J.E. (1964). Basic Marketing: A Managerial Approach, *Homewood.*

Miomir, M. J., Ljiljana, K., Miljan, J. and Vladimir K., (2017). Decision tree analysis of wine consumers' preferences: evidence from an emerging market, *British Food Journal*, Vol. 119 Issue: 6, pp.1349-1361.

Philip K. (2005). Principles of Marketing, *Prentice Hall.*

Ronald R. (2012). A View on Big Data and Its Relation to Information, *Chinese Journal of Library and Information Science*, Vol. 3, pp.12-26.

Treasuredata (2005). At Subaru, Data Puts Marketing in The Driver's Seat [Online]. Available at: https://www.treasuredata.com/customers/subaru/.

Vikor M. and Kenneth N.C. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think, *Eamon Dolan/Houghton Miffllin Harcourt*.

Wen, D.F, Zhu, H. and Zhu D.Q. (2015). IT to DT: Big Data and Precision Marketing.

Wu, Q.S. , Yu, Y. and Zheng, X.S.(2006). Thoughts and Methods of Precision Marketing, *Marketing Guide,* Vol. 5, pp.39-42.

Xu, H.L. (2006). Precision marketing theory, *Beijing Machinery Industry Press.*

Xue, L.Y. (2018). K-means on Application of Algorithm in Geological Hazard System.

Yin, Y.F. and Sun, C.Y. (2018). Application Research of Decision Tree Algorithm in Practice Teaching, *Computer and Digital Engineering,* Vol. 6, pp.1078-1082.

Zhang, J.M. and Gui, H.B. (2014). Application of Decision Tree Model in Employment Analysis of Higher Vocational Education, *Cyberspace Security,* Vol. 7, pp.81-84.

Zhang, X. (2014). Research on Decision Tree ID3 Classification Algorithm.