



Application of Data Mining in Business Analysis

Xinyue Li

*Business school, Kaiser University, West Palm Beach, Florid, American 33490
1251995934@qq.com*

Abstract

In the big data era, data are the future. Efficiently dig and analyze the potential business value behind a large amount of data is an urgent problem for marketing industry. Mining effective information requires the use of business analysis. This paper discussed the necessity of business analysis for the marketing, as well as some methods that can be used in making business decisions and data mining, finally discuss the future research and development fields about business analysis and data mining technology. When selecting analytical methods, it should be combined with the actual business situations and decision analysis purposes. When a certain method has defects, analyst can consider combining some different analysis methods to make a complement with each other. When data mining technology is applied to different social industries, it is necessary to notice the privacy and confidentiality of data information. Last, in addition to helping marketing managers find competitive advantages, business analysis can also combine different fields or industries through a strong data relationship network to optimize resource allocation and utilization, help decision makers to make efficient and valuable decisions.

Keywords-*Business Analysis, Data mining, Big Data, Marketing, Business Decision*

1. INTRODUCTION

With the development of science technology and the Internet, enterprise and individuals generate lots of data in daily life. The sources of a large amount of data are also widely, the data come from customer shopping records on the internet or game data on game platforms. According to statistics, the e-commerce website Amazon has more than 200 million active customers, all Facebook users can share 2.5 million pieces of content per minute.[3] Nowadays, companies want to have long-term development, enterprises need to use business analysis to dig the business value behind the data. For example, through the existing data to analyze customers consumption habits and behaviors, and conduct targeted marketing, because marketing is a system that composed of customers, companies, products and related environmental factors, which can show products in a certain country or region the relationship that best meets customer needs.[4] Also, forecast the future development trend of the company is convenient to make business strategies and improve business performance. This paper mainly discusses the application of business analysis in marketing.

Based on the analysis of the article, it is divided into the following parts: (1) Related concepts, types and processes of business analysis. (2) 5 methods that data mining technology can be used in marketing. (3) The future development direction about the data mining technology and business analysis.

1.1. The definition of business analysis

Business analysis starts with a simple data set or database, a database is a data collection that contains information such as customers and clients locations.[3] Business analysis needs a large amount of data, big data are simply defined: A collection is composed of many data with complex structure and diversified types.[9] Business analysis can be described to start from collect related data, the process is composed of descriptive analysis, predictive analysis and prescriptive analysis components, and the most important is the result of analysis is beneficial to make business decision and organizational performance.[2] Table 1 shows three types of analysis and definitions.[2]

TABLE 1. TYPES OF ANALYSIS [2]

Types of Analysis	Definition
Descriptive	Use simple statistical graphs to describe the contents of a data set or database
Predictive	1. Use statistical software, operations research methods to identify the variables to be predicted 2. Build models to discover relationships and trends between some variables
Prescriptive	1. Through decision science, management science, and operations research methods 2. Allocate resources reasonably and optimize resource utilization

Next, through a simple example to show descriptive analysis. Figure 1 can directly find that China’s per capita national income has increased from US\$7,940 in 2016 to

US\$10,390 in 2020 but the descriptive analysis do not reflect the reason and future trend of the increase in per capita national income.

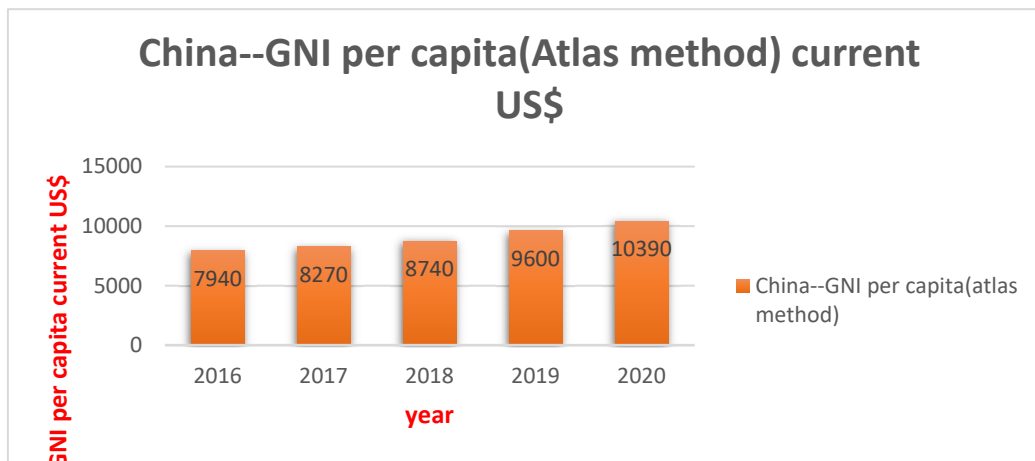


Figure 1 China---GNI per capita, Atlas method (current US\$)

(data from world bank) <https://data.worldbank.org/country/china>

Predictive analysis can use decision trees, multiple regression, data mining, ranking or predictive models, first find the relationship between variables, then through existing relationships to predict the likelihood of future occurrence of a thing. For example, food companies can research the relationship between customer’s age, weight, sports and weight-loss meal sales. Compared with descriptive analysis and predictive analysis, prescriptive analysis is a higher-level analysis, prescriptive analysis can use linear programming and decision theory,[2] (page 5) Also, optimization is one of the methods. For instance, use optimization can help retail industry finds a suitable inventory level, this inventory level can minimize product

shortages, retail businesses can meet the needs of customers at any time, corporate profits can be guaranteed.[1] (From the introduction)

1.2. Processes of business analysis

From the flowchart Figure 2, business analysis is divided into three stages. After obtaining the business data, the analyst needs to find out what information is included in the data that is helpful to solve the problem. Then, different problems need to be solved in the next three stages. Solving these problems requires the analyst to continuously dig out information from the collected data.

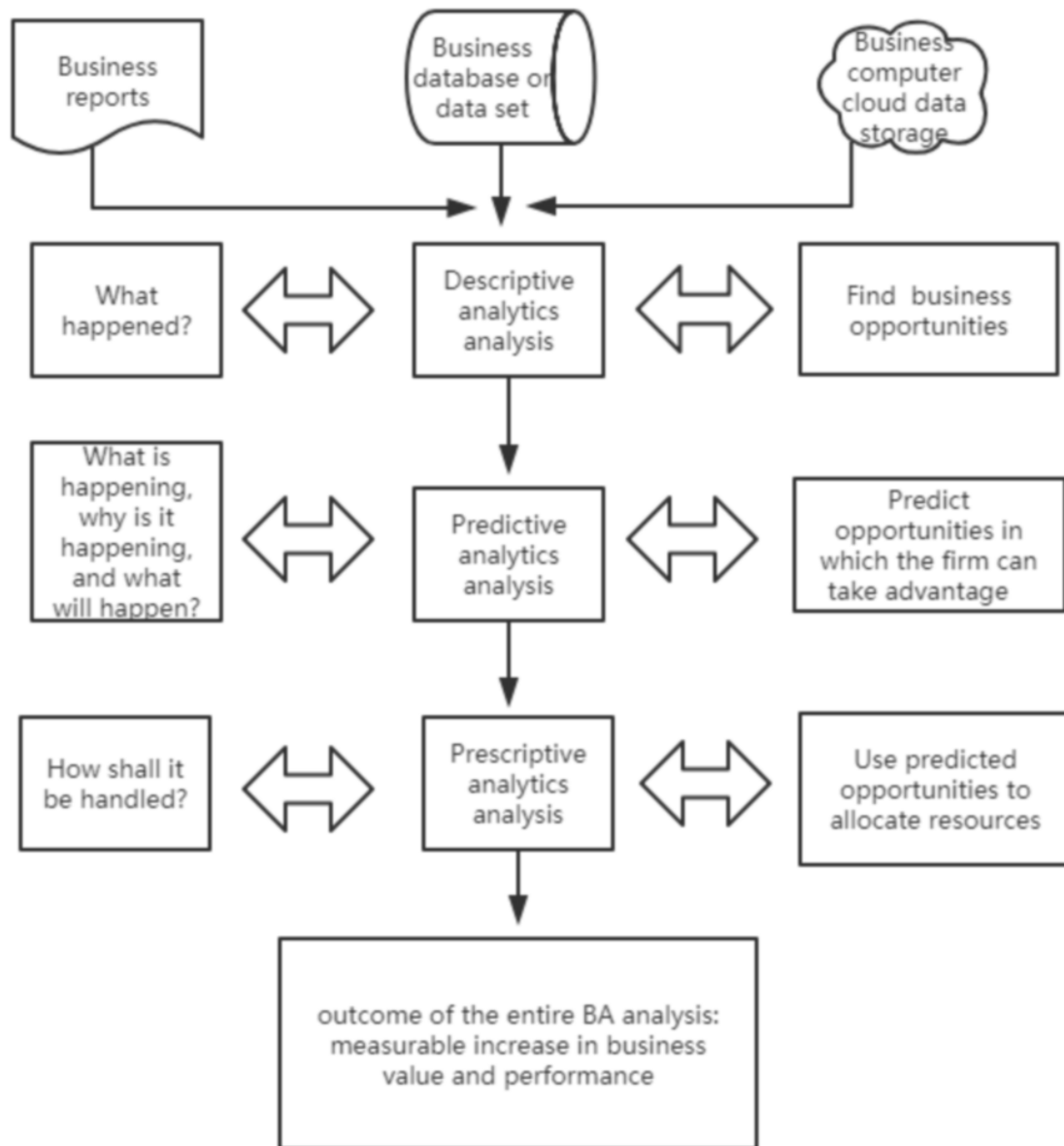


Figure 2 Business analytic process[2]

After collecting data through various channels, analysts will find some data sources are too complicated or difficult to handle, the data can be divided into structured data and unstructured data. Structured data are stored and managed in a relational database, or called database data. For example, the traffic data of a live broadcast platform. Unstructured data require analysts to manually process or use automated methods to edit. As an example, photos, sounds, words, etc. in social media. According to statistics, 80%-90% data obtained by many companies are unstructured data.[3] Business analysis first needs to handle the original data, convert unstructured data into structured data, retain valuable data. Figure 3 displays the steps of data process and analysis. In the first step, in order to decrease the amount of data, the

data can be averaged or use principal component analysis. For instance, When counting the geographic location of customers, it can be aggregated through spatial dimensions. In the second step, when the amount of data is large, Bayesian method provides a basis for data sampling. In the third step, analysts can select several models which are suitable for these data, then merge them to form a new model, which can overcome the shortcomings of a certain model. In the fourth step, the MapReduce algorithm divides a large amount of data into multiple data that can be processed in parallel, the final output result is the merged result of each segment.[3] Next, this research makes some charts to present the information displayed by the data more intuitively.



Figure 3 Data analysis and processing steps [3]

Later, in order to decide whether the trend presented in the descriptive analysis exists, whether the current trend can be used to predict the future development trend. Predictive analysis can predict which business opportunities previously discovered should be used. Such as promoting a certain lady's bag brand, what is the most suitable way of publicity. Finally, analysts communicate the results to the company's leaders and take corresponding actions, make decision by making use of the opportunities discovered after the forecast. Overall, the complete analysis process needs to be carried out around the business and clear the purpose of the analysis. Business analysis finds valuable information in data and forecasts the future trend through current data, helps managements to choose better business decisions, the most efficient allocation and use of resources to help companies improve business performance.

2. APPLICATION OF BUSINESS ANALYSIS

2.1. Data mining

Data mining is a key step in the KDD process (figure 4[5]), which can be defined as: using algorithms and statistical techniques to automatically or semi-automatically extract unknown and interesting patterns from large amounts of data.[1](From the introduction) Data mining has a wide range of applications, for example, retail industry, there are lots of data in retail industry, through analyzing customer shopping records and other information, analyst can distinguish customer categories, product categories and find out the most loyal customers, Then targeted publicity helps companies gain more potential customers, improve customer satisfaction, and reduce costs and increase revenue. The following introduces several methods of data mining.

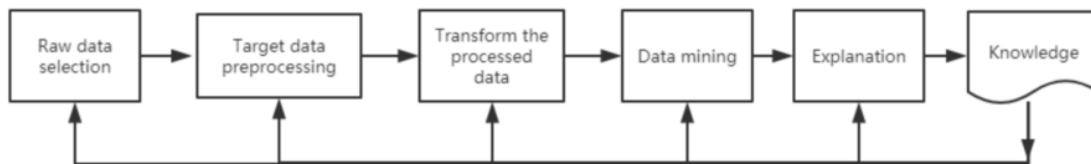


Figure 4 KDD process [5]

2.1.1. Cluster analysis

Cluster analysis of the original input set is not classified before, the purpose is divided into different categories based on different rules or characteristics.[5] For example, when doing market segmentation, according to the purchase behavior of customers, customers are divided into different groups, different customer groups have different promotion strategies. For the same input set, the groups divided by different algorithms are not the same. The following describes one of the algorithms---fuzzy clustering analysis. For some information that cannot be described by numbers, use fuzzy clustering models and methods can get better results, also, objects that are clustered into one category first have higher feature similarity.[8] For instance, if a company needs to classify n dealers, there are m evaluation indicators, the review team will score different dealers on a percentile system and average each indicator.

(Dealer) $X = \{x_1, x_2 \dots x_n\}$, (Index) $x_i = \{x_{i1}, x_{i2} \dots x_{im}\}$, generate a table with n rows and m columns, first, standardize the data in the table, the original data matrix is: $X = (x_{ij})_{n \times m}$, the formula is

$$x'_{ij} = \frac{x_{ij} - \min\{x_{ij}\}}{\max\{x_{ij}\} - \min\{x_{ij}\}}$$

Then use the arithmetic average minimum method to obtain the fuzzy similarity matrix R, use the square method to calculate the fuzzy equivalence relationship matrix R^* . Finally, the elements in R^* can be classified at any level, according to different λ values, form a dynamic cluster map, and get the required classification, classify the dealers with similar characteristics into one category, the company will have a better evaluation of the dealers.

2.1.2. Correlation Analysis

Correlation Analysis can be understood as finding the correlations hidden in big data sets. The application of correlation analysis includes planning warehouse inventory, network failure analysis, customer shopping analysis, etc.[5] Here is a brief description of correlation analysis to help enterprises understand customer performance, develop bundled sales of goods and products placed in position strategy. For example, a story about beer and diapers. The supermarket analyzed customer bills then found that beer and diapers, two seemingly unrelated items, often appeared in the same shopping basket. And later also found this situation happened to the young father, the supermarket increased profit through this discovery. Before understanding the steps of correlation analysis, introduce the four parameters of association analysis: ①The support degree describes the probability of two item sets (A and B) appearing at the same time, and the formula is $\text{Support}(A \rightarrow B) = P(A \cap B)$. Assuming that 50 people go to the supermarket to shop on Friday, and 10 people buy diapers \rightarrow beer at the same time, the support degree is 20%. ②Confidence degree indicates the probability of item set B also appearing in a certain situation where item set A appears. The formula is $\text{Confidence}(A \rightarrow B) = P(B | A)$. Assuming a customer purchases diapers, The customer is as likely to buy beer as it is to buy diapers. ③Expected confidence, if there are 50 people on Friday to go shopping, 10 people purchases beer, so the expected confidence is 20%. ④The degree of action refers to the ratio of the degree of confidence to the expected degree of confidence. The larger the ratio, the greater the influence of item set A on item set B.[7] When performing association analysis, first pre-process and mining data, after the database is formed, all the item sets satisfying the minimum support degree are obtained, then the rules satisfying the minimum confidence degree are generated, and finally the rule set is output.[5]

2.1.3. Classification analysis

Classification analysis is different from cluster analysis. The input set of classification analysis has been

classified and labeled according to different characteristics, according to the records, subdivided into new categories based on certain criteria.[5] Classification analysis includes bankruptcy prediction, and predict the reaction of different customers to promotional activities, etc.[4] For instance, bank credit rating for credit card customers, according to credit rating into good, fair, poor, and describe customers with "portraits", find different credit levels customers have different salaries, ages, bank deposits, and residential locations. Finally, banks can market according to different customer groups, launch different wealth management products, better help customers allocate funds, etc., so that customers are satisfied, and the bank can also obtain a group of highly loyal customers and achieve long-term development.

2.1.4. Decision tree

A decision tree is to use the information gain in information theory to find the field with the largest amount of information in the database, establish a unique root node, then establish multiple branches and nodes according to the different values of the fields, that is, leaf nodes.[5] Decision tree let people more understand, simple structure, high classification accuracy, and high efficiency.[6] Decision tree analysis method can be applied to customer churn in the market, managers can understand what factors affected customer churn. When making marketing strategies, managers can timely improvements to reduce customer churn rate and improve customer relationships. Assuming gets some customer records from a company database, these records can be divided into training samples and test samples (training set \rightarrow test set), then establish a classification prediction model and obtain a simple customer churn analysis decision tree, as shown in Figure 5. For high-value customers aged 25-30 with income greater than 8,000 yuan, the company should do a good job of customer service for such customers and reduce the rate of customer churn.

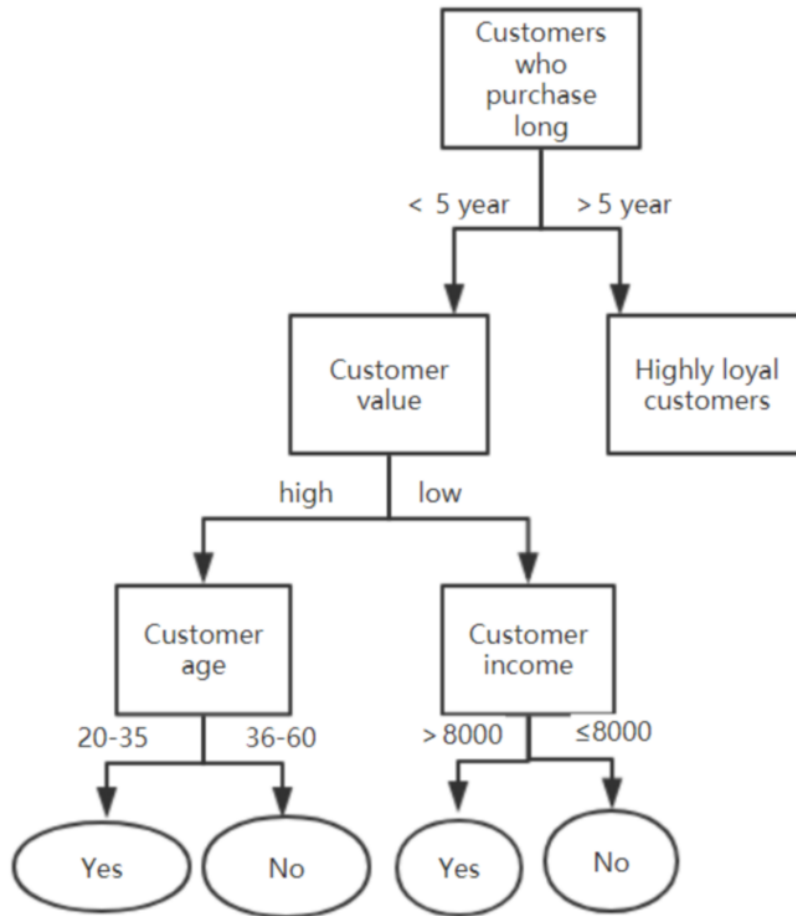


Figure 5 Customer Churn Decision Tree

2.1.5. Neural Networks

Neural networks have fast parallel computing power, and are a parallel information processing system composed of a large number of nonlinear processing units through dense connections.[4] Neural network error rate is low and a high tolerance for noise data, neural network can be used to predict the sales of a company's products at a specific time in the future, based on past sales, consider the impact of internal and external environment, to determine the future sales reachable the goal.

3. CONCLUSION

This paper talks about big data promote the development of marketing, help companies create more profits through business analysis, carding a business analysis of related concepts, types, and five kinds of data mining methods can be used, the purpose is to meet customer requirements, develop efficient business marketing decisions. When analysts use fuzzy clustering analysis to evaluate dealers, in order to ensure the objectivity of the results, it is necessary to increase or decrease the analysis indicators according to the different evaluation requirements of different companies in the actual situation, and draw more reliable and practical conclusions.[8] When analysts use association analysis

method, analysts should attention to three elements of the marketing theory: customer purchase time, purchase frequency and purchase profit, propose a mining algorithm with practical application value.[5] Although neural network can tolerant high data noise, when the amount of information is too large, the structure is complicated and the training time is too long, it can be considered in combination with the advantages of rough set (simple algorithm and easy to operate). Because in the reality marketing database, the data is large and the noise is high, two methods combine can maximize the advantages and avoid the disadvantages.[4] In general, marketing is a significant field of data mining technology. Data mining technology helps companies improve competitiveness, and gain better development. In addition, data mining technology is also involved in machine learning, artificial intelligence and other fields, they have high research value and application prospects, data mining will gradually be applied to all areas of human society in the future, but during the process of data mining should also pay attention to data privacy and information security issues.[7] In addition to the importance of business analysis to marketing, other social fields can also benefit from business analysis, such as medical and health, government departments, etc. The data relationship network connects marketing with other areas of society, so that enterprises and governments can share some data

resources, which is conducive to the rapid development of the national economy as a whole, optimizes the allocation of national resources, and achieves mutual benefit and win-win results.[9]

ACKNOWLEDGEMENT

I would like to thank the teacher who guided the thesis. Every time I make a revision, the teacher patiently put forward a lot of valuable suggestions. I would also like to thank my parents, who always encouraged me and gave me spiritual support when I encountered difficulties in writing the thesis.

REFERENCES

- [1] Davenport, T. H., & Kim, J. (2013). *Keeping Up with the Quants: Your Guide to Understanding and Using Analytics*. Harvard Business Review Press. (T. Translator, Trans.). Sheng Yangyan. (Original work published 2018) (From the introduction)
- [2] Schniederjans, M. J., Schniederjans, D. G., & Starkey, C. M. (2014). *Business Analytics Principles, Concepts, and Applications: What, Why, and How* (FT Press Operations Management) (1st ed.). Ft Pr. (Page 3-8)
- [3] Yang Yang, Liu Sheng, Li Yiwei & Jia Jianmin (2020). Big Data Marketing: Overview and Prospects. *System Engineering Theory and Practice*, 40(8), 2150-2158. <https://doi.org/10.12011/1000-6788-2020-1187-09>
- [4] Cui Huali. *Data mining method and its application based on rough neural network*. (Doctoral dissertation, Xi'an University of Architecture and Technology).
- [5] Chen Jinxiu. (2002). *The application of data mining in marketing*. (Doctoral dissertation, Xiamen University).
- [6] Sheng Zhaohan & Liu Bingxiang. (2005). Decision tree method for customer churn crisis analysis. *Journal of Management Science*, 8(2), 20-25.
- [7] Yu Chunhong. (2010). *Business analysis in the computer retail industry based on data mining technology*. (Doctoral dissertation, East China Normal University).
- [8] Qiang Fengjiao. (2006). *Application of Fuzzy Cluster Analysis in Dealer Selection Evaluation*. *Shopping Mall Modernization*, 000(001), 72-73.
- [9] Hu, Y. (2018). Marketing and Business Analysis in the Era of Big Data. *American Journal of Industrial and Business Management*, 08(07), 1747-1756. <https://doi.org/10.4236/ajibm.2018.87117>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

