Advances in Economics, Business and Management Research, volume 212
7th International Conference on Economy, Management, Law and Education (EMLE 2021)

Analysis on the Application of Data Science in Business Analytics

Yiran Wang[1,*]

[1] *University of California, Santa Barbara, Santa Barbara, California 93106, the United States*
[*] *Corresponding author. Email: yiranwang0810@gmail.com*

ABSTRACT
Business analytics is a new term, which can be seen as an expanded filed of data science. The mathematical formulas, statistical models, and programming skills in data science help companies to utilize big data and collect useful information from customers. However, for data science, it just collects numerical values from different resources, and uses programming tools to analyze data in the model to get numerical results. To convert numbers into useful information, people need to use business knowledge to interpret numerical results. With clustering models in data science, customers can be segmented into different groups based on their backgrounds including basic information which helps the company learn characteristics of each customer groups. Once machine learning is incorporated into business intelligence, the algorithms can use historical data as input to predict the trend of market and their customers' behaviors as well, helping company improve productivity, quality, and customer service and so on. Data science gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of companies' business in the future. It is more like a tool for business to help companies address problems in operations, provide better services for customers, and make their business long-lasting and profitable.

Keywords: *Machine learning, Business analytics, Data science, Segmentation, Customer value, Measures.*

1. INTRODUCTION

Business analytics is a relatively new term in today's world. It can be connected to mathematics, statistics, programming, and business. Data scientists use complicated mathematical tools, machine learning and programming methods to deal with a variety of data and provide expert knowledge based on business strategy. The mathematical knowledge and programming tools help business managers to make decisions in a shorter time and solve problems.

People are living an era of big data with high technologies. It is inevitable that everyone leaves their marks on internet, and their data, such as shopping history, will be recorded in various forms by internet enterprises for the future business purposes. The availability of large data sets in business is a valuable recourse for the companies, and those data can be used in different fields of management and decision making. If there are a large amount of data sets about consumers, data science is the most important knowledge people need to learn before they start doing analytics. Comparing with business analytics, data science is a broader term which helps people turn the numbers into real value. By analyzing those data, people can predict the shopping decisions of the consumers, classifying their customers, and sending out advertisements to consumers who have demands on those products.

Through the raw data from customers' side, the consumer behaviors can be seen through the raw data, such as shopping history, past feedbacks for old products, and communication records with after sale services. With the enough data, analytics are used to convert data into useful information to companies. With consumers' shopping records, the company can see which type of products they purchase the most, how often they purchase the products, and how they rate the products and other consumers' behavior [1]. To convert number into

useful information, data scientists need to use machine learning models and programming tools to analyze the raw datasets. The models can help them to predict the consumers' next move on purchasing products under certain situations. Also, the data scientists can cluster consumers based on their living standards, income, and some personal information. Let's take a clothing brand as an example. If the data scientists from that clothing brand have some data of their customers, they can divide customers into different groups by ages and income. With specific groups of their customers, the company can send different advertisements to different groups and make products based on each group's demands. This helps a company to develop their business strategies and expand markets.

The main topic of this paper is to find the role of data science in business analytics and how do companies benefit from application of data science. By looking through different datasets, such as customer survey results of a supermarket and transaction history of a shopping websites, the author applies different machine learning algorithms on these datasets to see how a business use segmentation to gather information of their customers to improve their service qualities. Clustering methods such as k-means, and decision trees are used for classification purposes. Based on the results of data analysis, it can be seen how data science impact business in different aspects.

2. APPLICATIONS OF DATA SCIENCE

2.1 Market Segmentation by Data Science

When a company wants to sell products to consumers, it is necessary to segment is customers into distinct groups, and each group has its unique natures. Through the segmentation, the company can tell how each customer group is different from each other. If a company just collects basic information from their customers, such as job, income, age, there can be enough information for the company to segment their customers into different groups. Customers with different age and different income can be put into more than eight groups: the old people with high income, young people with low income, and young people with high income. This is just a simple example of how the customers are classified based on the basic customers' data. Imagine there is a customer group, the average age of the customers is 25, and they have comparatively low income. Most of them are working classes who just started working a few years ago. To those customers, there is potential purchasing power on them, although their incomes are comparatively low during this period. It is possible they will have higher income in the future, and they will be able to afford products with higher prices. It is important for the company to keep connections with those customers [2]. Based on the information converted from data, the company can choose to send out monthly advertisement of products to those customers, and pick the products which are affordable to them and satisfy young people's demands. Also, based on those customers' income, the company can think about sale strategies for those customers to improve their potential purchasing power. The company can give discounts only to this group of customers, and cover some costs on shipping to simulating their potentials on consuming. There are a lot of ways on how the consumers' data can be interpreted, and the business decisions can be made based on analytics to improve company's profits gained from different group of customers [3].

2.2 Customer Value Measures

The knowledge on data science can help us to measure the customer value in business analytics. By knowing customer value, companies can aim to improve the productivity of marketing activity and increase profits of business by identifying the value of different customer. Also, the customer value can motivate companies to plan their business well and manage resources to maximize their profits. The customer value measures make companies understand which customers bring values and give different treatments to those customers. They can also target the customers who have great customer value, then provide unique products and services to them, such as those VIP customers of some airline companies. Those VIP customers take the flights very often, and usually only take flights of one company. Since they often spend a lot of money on traveling, the airline company values the profits those VIP customers bring to the company, and they need to think about how to keep the loyalty of those customers. Therefore, the term "VIP" comes out. The airline company provides special service to the customers they value the most. Before the airplane arrives, VIP customers usually can take a rest in a comfortable room and enjoy meals provided by the airline company. Once the airplane arrives, they can get on the airplane before everyone else [4]. Also, for those customers, the company always thinks about how to provide great

service to satisfy them and let them be willing to spend more money. Thus, those rich customers who travel a lot are always targeted by the airline company. They will be served additional service and great food on airplanes, and other normal customers can not enjoy the same level services as VIP customers do.

Today, people have so many options of purchasing one specific kind of products. The online shopping and efficient delivery allow people to go through more brands and more products when they have demands on a specific kind of things. Therefore, from the perspective of companies, it is important to find out the customers who bring values, and the companies need to value those customers' satisfaction and loyalty to their brands.

3. ANALYSIS METHODS

3.1 Segmentation Methods

From the data scientist' perspectives, there are many good candidate models to help with segmentation: K-mean and decision tree. Clustering is a common technique to analyze data and get a basic learning about the structure of the data. Clustering divides data into subgroups and data points in different clusters have different features. The data points in the same cluster are similar. To do market segmentation, clustering is the most important technique. This is considered as an unsupervised learning method: finding natural groupings of data points in independent variable space.

K-means clustering is a method of vector quantization, which aims to partition observations into k clusters in which each observation belongs to the cluster with the nearest mean. K-means clustering minimizes within-cluster variances. Each data point from the data set only belongs to one group. The number of clusters K needs to be specified. Initialize centroids by first shuffling the dataset and then randomly selecting data points for the centroids without replacement, and keep iterating. To explain k-mean clustering in a simpler way, the algorithms just used stance-based measurements to determine similarity of data points, and put them into different clusters [6].

Another clustering method in data science is decisions tree. A decision tree divides the independent variable space into boxes and places a step above each box at the height of the average of the dependent variables in the box. The decision tree model carves up space into boxes and then assigns score for each single box. If the segments are clear, a big rectangle divided into different small boxes, and each of them has different natures. Segmentation divides customers into separate distinct groups, and each customer belongs to one segment with different characteristics from the other segments. Based on different characteristics of each segment, the business can make different decisions and take different actions for each different customer segment [5]. The business can offer different product, give discounts to specific groups, provide different services, and send different marketing messaging.

3.2 CVM Methods

Generally, customer value measures (CVM) are a measurement formula, combining known data values with some models. For some simple formulas, it can be simply related to linear regression in data science. There is a basic form for customer value measures.

CVM = profit from each customer = revenue - cost

In a straight way, customer value is the profit that the specific customer brings to the company. It usually is the amount of money a customer pays for the products, and minus the cost of making the products. Thus, the data used to analyze customer value measurements can be simply gathered from transaction history, and the company can calculate the costs by themselves [7].

Sometimes, the company needs to consider other factors which might influence customer value measures. The previous formula we talked about only shows the customer value from the direct profits company gains from transactions with customers. However, companies also need to consider loyalty when they measure the customer value. Some customers might not purchase products from one company very often, but they might always keep the company as their first choice when they need to purchase certain items. The best examples are big fans of technology companies, such as apple. Every time apple released new phones, some old fans might not purchase the recently released model, but once they need to get a new phone, they will always consider apple as their first choice. Such loyalty is one of the most important factors companies need to conclude in their customer value measures. If a company wants

to build their brands and diversify brand cultures, it needs to learn how to value customers' loyalty and how to provide best service with those loyal customers. It cannot be simply used with linear functions to analyze CVM to consider loyalty in customer value measures. Polynomial function will be a good candidate.

In addition, the company can use customer value to better understanding the results of customer segmentation. They can have a basic learning that what do high value customers look like, and how are they different from low value customers. The characteristics of different customer segmentation can help them learn to relate customer values to the customer backgrounds. After that, it will be easier for them to do targeted discounts, and optimization of particular business actions [8].

4. CONCLUSION

Consumers in today's world have more opportunities of learning new brands, and they have more access to new products by shopping online. Thus, the competition between brands and companies is intense, and it is more difficult for small companies gain supports from consumers. Utilizing data science is a good way to prepare companies in business war. Gathering consumers' data, providing good service, targeting loyal customers…such methods and actions are based on business analytics, which converts numerical values of data into useful information to companies. First, decision tree is an efficient way to do segmentation of customers for marketing purposes based on historical data, but the segmentation results still need a business expert use his judgment and experience to interpret and make wise decisions to help companies improve their service and the quality of products. Second, customer value measures (CVM) is an important tool for a company to have a comprehensive understanding of their business targets and the choices on formulas can satisfy different demands, but behavioral data of customers are often difficult to obtain, which could be an issue for small companies.

The advent of data science in the globally competitive business industry has positive impacts, and it plays as a new decision-making tool for the business analysts to deal with complicated situations in commercial cases. It is not a short-term trend for applying data science in the business world where abundant raw data waiting to be converted into actionable insights for making better decisions. In the future, doing analytics based on data science for business purpose will become inevitable in most companies, and more people will learn that data science is inherently linked with business analytics.

AUTHORS' CONTRIBUTIONS

This paper is independently completed by Yiran Wang.

ACKNOWLEDGMENTS

Throughout the learning of business analytics, I received support and assistance from professor Coggeshall and teacher assistant Yuhui Li. I used slides from lectures and python notebooks from class as references for part of the contents.

REFERENCES

[1] Blocker, Christopher. Modeling customer value perceptions in cross-cultural business markets [J] Journal of Business Research, vol 64, Issue 5, 2011, pp 533-540.

[2] Gallarza, Martina & Gil-Saura, Irene & Holbrook, Morris. The value of value: Further excursions on the meaning and role of customer value, 2011.

[3] Lieder, Michael & Asif, Farazee & Rashid, Amir. A choice behavior experiment with circularbusiness models using machine learning and simulation modeling [J] Journal of Cleaner Production, vol 258, 2020.

[4] Bichler, M., Heinzl, A. & van der Aalst, W. M. P. Business Analytics and Data Science: Once Again? Bus Inf Syst Eng 59, 77–79 (2017).

[5] Floh, Arne & Zauner, Alexander & Koller, Monika & Rusch, Thomas, Customer segmentation using unobserved heterogeneity in the perceived-value–loyalty–intentions link [J] Journal of Business Research, vol 67, 2014, pp 974-982.

[6] Earley, Seth. Analytics, Machine Learning and the Internet of Things [J] IT Professional, vol 17, 2015, pp. 10-13.

[7] Chen, Yubo & Leung, Carson & Li, Hao & Shang, Siyuan & Zhang, Wanmeng & Zheng, Zhi , A Data Science Solution for Supporting Social and Economic Analysis [J] Annual

Computers Software and Applications Conference, 2021, pp. 1689-1694.

[8] Halibas, A. S. & Shaffi, A. S. & Mohamed, Mohamed, Application of text classification and clustering of Twitter data for business analytics [J] Majan International Conference (MIC), 2018, pp. 1-7.