Consumer Behavior Analysis from Bigdata Insights: Evidence from YouTube, Douyin and Bilibili

Haoran Li

College of liberal arts, University of Minnesota Twin Cities, Minneapolis, the U.S

lixx5506@umn.edu

Abstract. Some businesses, in the age of intense rivalry and rapid expansion of big data, have started paying attention to the actions of their customers, using big data analysis in order to evaluate and keep tabs on their customers. By evaluating user activity using big data, businesses are better able to comprehend the requirements of customers and provide them with more individualized offerings. This article estimates the instances of YouTube, Douyin, and Bilibili in order to highlight the impact of big data on consumer behavior as well as the use of big data in the study of consumer behavior. In the area of OVP, three separate businesses used a variety of approaches to big data processing in order to carry out in-depth consumer research, and the outcomes of this study were all deemed acceptable. In conclusion, this article discusses some of the current limitations of big data as well as the prospects for the future, with the goal of helping as well as various insights to businesses if they decide to conduct in-depth research on customers based on big data technology in the future.

Keywords: Consumer Behavior Analysis, Online Video Platform (OVP), Big Data.

1. Introduction

When productivity exceeds market demand and competition is intensifying today, some companies begin to pay attention to consumer behavior. The study of consumer behavior is a relatively new branch of social science that first made its debut in Western publications in the middle of the 1960s. On the other hand, it is possible to trace the origins of consumer behavior all the way back to the beginning of human consumption and life. Research on consumer behavior is being conducted in a growing number of academic disciplines, including economics, sociology, and psychology [1]. Contemporarily, there are billions of individuals who consider regular use of the internet, various social media platforms, mobile apps, and other forms of digital communication technology to be a normal part of their lives [2]. Communication over the Internet in particular, a new form of consumer socialization, has far-reaching implications. Influence consumer decision-making, and then influence marketing strategy [3]. Research on consumer consumption behavior is the behavior of the company's evaluation of consumer behavior. When understanding current trends and consumer needs, it is helpful for companies to formulate marketing strategies and provide them with market advantages [1]. It is the significance and benefits of studying consumer behavior.

Although there are numerous technologies available today that are capable of doing consumer behavior analysis, the market for such services is still relatively undeveloped and has a significant amount of unrealized potential. Big data is one of the most developing technology developments, and it has the astonishing power to revolutionize corporate organizations by leveraging customer behavior to study and convert them into valued customers [4]. This ability gives big data its name. In the traditional offline world, a single commodity makes it difficult for consumers to find substitutes. The problem in the digital world is that there are so many choices, a dizzying array of items to sift through. Hofacker et al. mentions that direct-to-consumer retailers would log all purchases in the past, but now retailers can log all search activity purchases on their website and shopping apps as well as a log of all activity on the website, which items are included has been purchased, search clicked, added to cart, etc. [5]. People want a breakthrough in personalization and real-time content delivery to social network users. This requires efficient, event-based algorithmic structures for big data processing, characterized by finding approximate elements in dispersed, dynamic information. Then, one marks

the elements of interest and present them to users, achieving the goal of giving users preference in online marketing strategies. Currently, text search and indexing technologies are relatively mature. However, video and image mining technologies are still at a relatively early stage. This new challenge requires powerful offline automatic image and video indexing and searching, which brings additional resistance to the development of big data analysis [6].

People's consumption patterns have undergone significant shifts, which has led to the emergence of and rapid growth in online consumption. These shifts have been brought about by the continuous development of Internet technology, the continuous improvement of social and economic levels, and the popularization of various communication devices and mobile tools. Compared with traditional offline platforms, online platforms are more convenient and have more options. In the process of using social networks, users generally perform operations (e.g., searching, viewing, and shopping), thereby generating a large amount of data related to consumers. The effective usage of these data requires the use of technologies and tools related to big data. Big data technology collects data on user habits and styles to effectively analyze users and formulate different platform marketing and personalized services. The network platform is based on sophisticated data mining technology, and the marketing of the platform is no longer limited to conventional mass marketing. Instead, the marketing of the platform focuses on personalized marketing activities that are based on large amounts of data. Under the big data environment, the network platform mainly exhibits the characteristics of satisfying the personalized needs of consumers. Secondly, when users conduct online activities at home, the network platform can analyze the accompanying needs according to the user's past activity habits, to analyze user portraits, infer preferred behaviors, and carry out targeted marketing. The traditional consumption model is based on the offline form, which often leads to a lot of time consumption in the selection. The network platform enables the integration of resources and can accurately meet the needs of consumers according to the habits and behaviors of users, and serve consumers Provides great convenience. The research purpose of this article is to analyze how three online video platform companies YouTube, TikTok and Bilibili use big data to analyze consumer behavior in the big data environment, as well as the results achieved and the help to the company's development. Afterwards, this paper will also discuss the limitations of big data analysis of consumer behavior, as well as the outlook for the future.

2. Basic descriptions of bigdata

The progression of big data technology will have a substantial effect on the application of that technology to the study of consumer behavior. Maturity as well as the Internet's rapid development in information technology. With the use of appropriate technical methods, big data technology can gather a significant amount of consumer data in order to analyze consumers' purchasing intentions. Big data analyzes consumer data related to search, browsing, and behaviors in order to ascertain whether consumers are willing to make purchases. It then executes accurate product delivery to consumers in order to fulfill the goal of encouraging consumer consumption.

The average customer of today is a constant producer of behavioral data as well as more modern transactional data that is both structured and unstructured thanks to technology. Marketing decision-making is changing as a result of the production of sheer volume of data, the dizzying rate at which new data are created all the time, as well as the available enormous variety of data. Volume, Velocity, and Variety, also known as the three Vs, are three characteristics that help define big data [7]. In the lead is volume. Big data's primary distinguishing characteristic is volume. The need for software, hardware, and services for storing and processing big data is growing steadily on a worldwide scale due to the information's accelerating expansion. Big data's fundamental performance is volume.

The second key to big data is Velocity, which is followed by Velocity. The concept of "Velocity" refers to the ability of marketing executives to make better judgments based on time-specific evidence as opposed to intuition or lab-based consumer research when they have access to rich, insightful, current data. Big data allows its marketing director to determine how many consumer visits and

transactions take place over a certain time period as well as which items customers most commonly buy and visit in a constrained amount of time. Executives in marketing utilize this information to make recent, fact-based judgments.

Variety is the final crucial phrase. Rich and diverse material is provided by many big data sources. The most notable difference between conventional data and current big data is the shift from structured transactional data to unstructured behavioral data. This is an important step in the evolution of big data. Marketers gather structured data, such as information from scanners or sensors, documents, files, and databases. Present-day big data is mostly comprised of unstructured data, include text data (such as messages) and non-text data (such as images or videos) [7]. The online video platform (OVP) of today collects a vast quantity of unstructured data via social media, which is where consumers communicate personal and behavioral information with friends and family. With this information, they may acquire a better understanding of consumer behavior preferences.

The majority of data mining and statistical analysis tools are often created with huge datasets in mind. In reality, many users develop or manually write sophisticated SQL that parses massive data to discover the suitable customer base rather than employing mining and statistical tools, which leads to churn or unnecessary operational expenses [8]. Big data management tools are many at this time and a significant amount of unstructured data is being created in numerous formats, including audio, video, pictures, text, and animation in the current big data era. The capacity of Hadoop to scale to big data quantities is astonishing. Hadoop manages analysis and query preferences across various data kinds using MapReduce, and it works well in current unstructured data processing. Hadoop's file system is able to accommodate a wide range of data formats because to the software's flexibility.

There are already a wide range of choices for big data analytics programs. Options include user technology and approach, team or organizational structure, vendor tool type and tool capabilities. The optimal course of action always be chosen, and software or hardware projects must be constantly assessed, thus being aware of the possibilities is essential. In order to quantify these and other challenges, TDWI offered survey participants a wide range of big data analysis choices [8]. The list covers solutions that are have been proposed earlier (cloud, MapReduce, complicated event processing) but just now gaining traction (data visualization, predictive analytics), or are well-established and have been around for a long time (statistical analysis, hand-coded SQL).

3. Youtube

YouTube is a website that allows users to download, view, and share videos or brief snippets. The original firm was founded in San Bruno, California, in the United States, on February 15, 2005. Traditional YouTube data collecting and analysis Frequently used techniques for gathering information on YouTube viewers include interviews, focus groups, surveys, content analysis, and experiments [9]. This paper often refers to this kind of study as social media analysis when it comes to data produced by social media sites (SMA).

These analytics provide insightful data on the opinions, behaviors, and activities of internet users. Data from the web and social media have been studied using a variety of social science research methods now in use. SMA may also be seen as a collection of methods for deriving significance from data gathered from social media platforms and assessing the effectiveness of social activity. The YouTube application programming interface (API) and the YouTube interface may both be used by researchers to manually acquire YouTube data. APIs for YouTube. On YouTube, automated data collecting often calls for some programming knowledge, either in R or Python. These data must be further preserved and cleansed as they are often enormous and might be considered big data before being thoroughly analyzed. Researchers using the YouTube API with no programming experience might depend on expert services and businesses that provide data from social networks like YouTube. Applications may interact and communicate with one another thanks to an API, which is a collection of protocols and standards. There are two different sorts of interfaces that may be used to understand how YouTube collects its data: APIs and user interfaces (UI). When YouTube users post videos to

their channels, they often want to know who is watching the videos, which videos are doing well, and where the viewers are located as shown in Fig. 1.

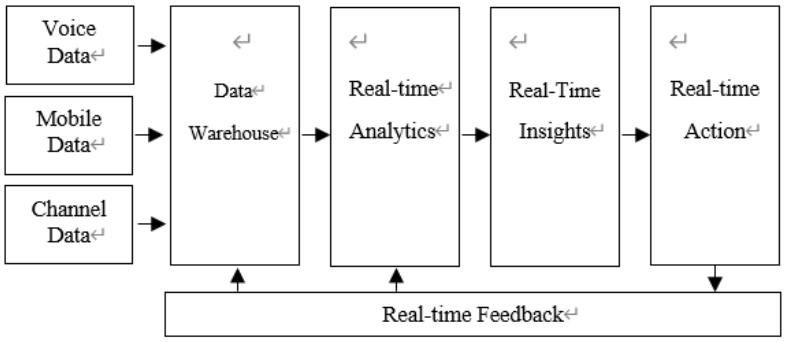

Figure 1. Big data process for customer Analytics

All the data is shown on a robust dashboard that YouTube Analytics offers along with several indicators. Similar to the Google Analytics architecture, YouTube Analytics offers insightful data on the website's video content. For a better knowledge of video performance, one ought to consider use YouTube video stats. YouTube analytics gives companies operating in a highly competitive climate the knowledge required to comprehend the reach more fully and viewability of their content. Measurement is crucial, and it allows consumers to keep tabs on their progress to see whether a certain objective is being accomplished. Consumer behavior and lifestyle preferences are some of the factors that YouTube may use to analyze client data. consumer behavior (e.g., how often they shop online) and lifestyle factors (e.g., their interests and hobbies). Everywhere that an organization stores its data, it must keep track of its consumers across all channels [10]. These objectives, which might include video views, pertinent involvement, etc., have excelled at user profiling.

4. Douyin

The debut date of the Douyin platform was September 20, 2016. Under "ByteDance," is a music-creative short video platform for people of all ages. It is a platform for sharing brief videos that were created to aid people in expressing themselves and documenting a better life. Traffic is advantageous in the Internet era since the audience's attention is widely scattered. In order to fully engage the audience and capture their attention, Douyin is dedicated to producing engaging short films that last between 15 and 60 seconds and combine audio, video, graphics, and text in novel ways. Additionally, Douyin has a very low production barrier, enabling users to make their own little video works without specialized professional expertise and share lovely moments with others whenever and wherever they choose. Users might be performers or audiences. There are no obvious limits or thresholds [11]. Based on artificial intelligence to provide engaging and varied gaming for consumers, they may quickly and simply make high-quality short videos in their daily lives. Douyin is a platform for filming and sharing short videos of life. UGC (User Generated Material), PUGC (Professional User Generated Content), and PGC (Professionally Generated Content) are the three primary categories of video content produced on the Douyin platform (Professional Generated Content). User-generated content (UGC) is the term used to describe the videos that regular Douyin users, who make up the great bulk of UGC, contribute [11]. After being reviewed for infractions by the system algorithm, videos uploaded by UGC users are then forwarded to online users by identifying keywords, structured and unstructured data, combined with user behavior preferences, and then superimposed recommendations based on user feedback and scrolling continues to recommend (seen from Fig. 2).

Short videos on Douyin are distributed using keywords, and the site carefully chooses the user's keywords to do so. Everyone who is interested will be chosen for the personalized content push. The information flow algorithm is the technique used by Douyin. The term "information flow algorithm" refers to the information flow rules that employ artificial intelligence to gather information about a

user's network activity, determine the user's interests using a specific algorithmic logic, and then provide the user with recommendations that are specifically tailored to their needs [12].

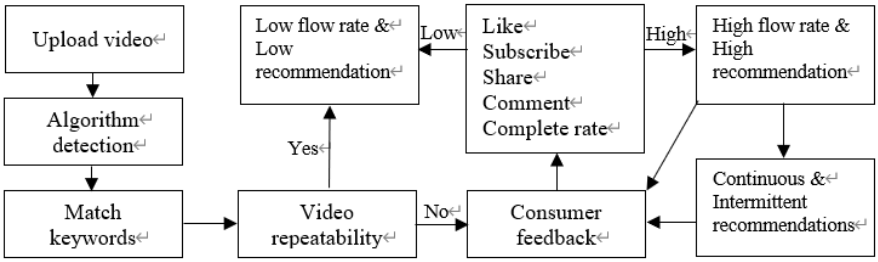

Figure 2. Videos upload algorithm process

Douyin's brief video analyzes user psychology, examines users' online activity patterns, and notes their preferences and routines. It continually draws users' attention with the display of tailored data and promotes videos that consumers want and need. Users' demands are satisfied, encouraging them to interact with Douyin more actively and view brief videos. The Douyin platform has a sizable user base, which has tremendous economic potential. Despite the fact that user data is dispersed, big data techniques have helped make the Douyin user group's interest labels highly distinct [12]. A similar user picture analysis is achieved by Douyin in order to optimize push impact and correct delivery. Building user portraits is the primary method used to place huge data precisely. The exact placement of people is eventually achieved, and the short video dissemination and conversion efficiency are enhanced, thanks to scientific statistical methodologies, user network behavior data, and keyword extraction. Second, real-time monitoring can be done in the background when short videos are sent via Douyin, and the distribution method and frequency can be changed over time based on how users react to the videos. Based on the reports that big data analysis has examined, they may also choose possible user groups for a push in a timely way. This feedback method, which is based on impact assessment and prompt modification, may successfully reach the target consumer groups the first time, prevent delivery that is done in the dark, and reduce delivery costs. For Douyin's large user base, the function of the exact placement of brief video transmission is essential. It is simpler to increase user satisfaction with accurate push since it is more in accordance with user demands and preferences. Additionally, it benefits the platform's growth. enhancement. The size of the short video market in China as of 2020 was 140.83 billion yuan. According to the survey's findings, Douyin is the most popular product among users of short videos, with a share of 45.2%. The number of active users has also climbed, from 0.55% in January 2018 to 490 million in November 2020 [11]. Through the application of this technique of big data analysis of user behavior, Douyin has moreover emerged as a pioneer in China's short video market.

5. Bilibili

Bilibili, which was established as OVP in June 2009 and reflects Chinese young culture, is another illustration of excellent big data analysis. When compared to YouTube, the Bilibili is a bullet screen online video platform, and its primary push material is essentially ACG (Animation, Comic, Game). Bilibili has developed a collection of interactive forms and an environment with a unique aesthetic over a significant amount of time.

One might now categorize Bilibili as an ACG-related new media platform. Bilibili has profited from the quick growth of mobile terminals and big data applications. As of December 2018, Bilibili's average online penetration rate and DAU (daily active users) were 8% and 17.9 million respectively, making it the fourth largest online video site after Tencent Video, iQiyi, and Youku [13]. In contrast to Douyin, the output user group for Bilibili is primarily PUGC, also known as Uploader. Users and Uploader realize the link on the information dissemination chain in the service area made up of suggested videos. Individuals are able to locate the information source for their selected material since the videos uploaded by Uploader are correctly pushed to certain user groups. Through the

employment of this information processing technique, people are less under pressure to filter information, their platform stickiness is increased. They may surround themselves with their chosen material, and platform content is delivered effectively [14]. The viewer may naturally see relevant suggestions beneath the video's keywords after clicking it. The vast majority of these suggested videos include subjects, artists, and material that are comparable to the interface that the user clicked. Users will often be drawn to the pertinent suggestions that are shown below and may be seen with ease. Differently, there are additional suggested videos that, although not necessarily similar to the video the viewer first accessed, are connected to it in some ways under the keywords of the recommended video, which is a collection of algorithms based on views, comments, likes, collections, and coins (a reward system for people who share good video). By creating top-notch video, the uploader acquires the necessary push mechanism, and then Bilibili pushes hot places (as exhibited in Fig 3).

In conclusion, the video recommendation material on Bilibili is created using a dual method of user collaborative filtering and content collaborative filtering, and consumers progressively veer away from the subject they were initially interested in while viewing videos [14]. In contrast to YouTube and Douyin, Bilibili is more likely to alter user preferences for behavior and ACG culture. It gives businesses a means to extend consumer use time. Users, however, could find it to some degree. It is restricted to the category of suggested videos and eventually veers away from the initial subject matter. For Bilibili, however, an effective strategy to increase user stickiness is to leverage the push mechanism using big data analysis.

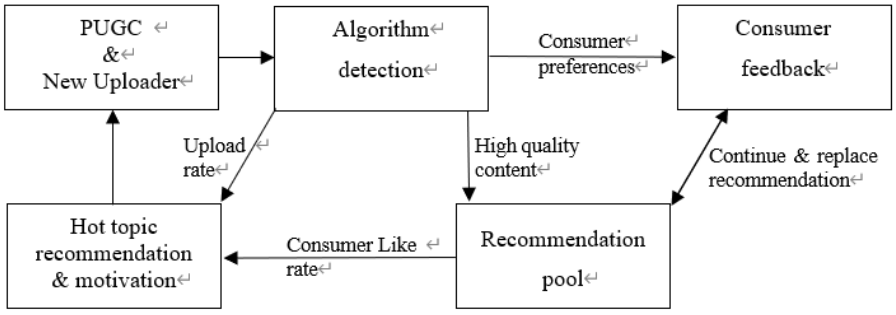

Figure 3. Bilibili video push mechanism

6. Limitations & Prospects

Although authenticity is the foundation of all data value, big data has a significant inherent flaw in this regard. One of the most significant sources of big data is the Internet, yet the Internet is also rife with incorrect information. The properties of the network itself govern this kind of distortion. For instance, it is challenging or even impossible for the majority of social networking sites to carry out a thorough verification of the veracity of member registration information. Additionally, the platform is powerless to prevent one person from creating multiple accounts or to regulate the connection between accounts and actual consumption. Authenticity with huge data is surely difficult to pursue. Big data itself, as well as its gathering and utilization, are likely to pose issues with national information security, trade secrets, and individual and public privacy. Therefore, one of the constraints limiting the utility of big data is security and legal concerns. Technology and business are vital, but the ideals that drive them are more crucial. Big data's low-density value characteristics mean that in order to unlock its potential, a significant scale of data gathering, and efficient value extraction are required. Useful information may only be included in a few seconds of the 24 hours of a day's website text records. Big data still faces a lot of challenges, including how to swiftly finish data processing and value display using sophisticated machine capabilities. The value of their data cannot be completely revealed due to the fact that many businesses and organizations' data cannot grow to the necessary levels according to the rule of rising marginal utility of big data.

Even while most businesses do Big Data-related tasks internally, some businesses outsource Big Data consumer analysis [7]. To better understand customer behavioral preferences, big data research

should be based on internal corporate study. Not all of the data in the vast quantity of data is helpful, and sometimes just a tiny portion of the data is useful. Meaningless redundant data and garbage data will proliferate as data volume rises, and their rate of expansion will outpace that of data information. Utilize other researchers' ignorance. Researchers are urged to put these ideas into practice and create theories that can be empirically tested. Therefore, while using big data insights, researchers are urged to create acceptable measurements of ignorance. When obtaining data from multiple data formats, this DSLR has a number of quality problems. These challenges may be broken down into a number of distinct areas, including data quality, data separability, scalability, diversity, and modeling complexity [15]. Inductive reasoning may be utilized to examine massive data using the principle of ignorance, modern technology, and cutting-edge algorithms. Researchers are urged to look at how the big data revolution and ignorance-based viewpoints may contribute to a potential rethinking of marketing research methodologies.

Using physiological data to understand customers more fully More forms of big data may now be collected by businesses owing to technological advancements (e.g., wearable multi-sensors). Some companies have recently begun to implement the use of sensing devices in order to gather, store, and evaluate biological data in real time, such as a person's temperature, respiratory rate, and neural activity [7]. The ability to glean a multitude of insights is provided by the combination of such real-time physiological data with behavioral assessments. Insights from big data, according to the researchers, may be used to analyze behavior and better understand behavior by using physiological measures. Researchers are therefore urged to consider the usefulness of physiological data in comprehending customer behavior. Accessing to Big Data from Consumers Big data has the potential to be a powerful new capital, and access to it gives a definite new approach for marketers to distinguish their offerings. A business has to draw a lot of customers to its product or service in order to access consumer big data. Some businesses choose an open-source approach over a closed-source one to accomplish this purpose [7]. Data collaboration and sharing are on the rise, with the goal of improving consumer big data access.

7. Conclusion

In summary, this study investigates the implementations of big data analysis in the area of OVP to conduct an analysis of customer behavior. For the sake of this discussion on big data analysis, this paper focuses on YouTube, Douyin, and Bilibili, respectively. According to the findings, mining and making use of large amounts of data has a significant influence on the study of consumer behavior, and this has a positive effect overall. In addition, the application of big data to consumer behavior makes it easier for businesses to gain a firm grasp on customer portraits. This makes it possible for businesses to ascertain the consumption preferences and potential consumption possibilities of consumers, and then to encourage those preferences and possibilities. In addition, the business should modify its marketing approach using the insights gained from big data research in order to get higher economic rewards. Nevertheless, there are still significant algorithmic and technological constraints associated with large data analysis. Big data analysis will continue to be a focus of interest in the area of OVP in the years to come. In general, these findings outline best practices for OVP insights to follow when doing research on customer behavior with big data.

References

[1] Oszust K, Stecko J. Theoretical aspects of consumer behaviour together with an analysis of trends in modern consumer behaviour [J]. Modern Management Review, 2020, 25(4): 113-121.

[2] Stephen A T. The role of digital and social media marketing in consumer behavior [J]. Current opinión in Psychology, 2016, 10: 17-21.

[3] Vinerean S, Cetina I, Dumitrescu L, et al. The effects of social media marketing on online consumer behavior [J]. International Journal of Business and Management, 2013, 8(14): 66.

[4] Khade A A. Performing customer behavior analysis using big data analytics [J]. Procedia computer science, 2016, 79: 986-992.

[5] Hofacker C F, Malthouse E C, Sultan F. Big data and consumer behavior: Imminent opportunities [J]. Journal of consumer marketing, 2016.

[6] Kambatla K, Kollias G, Kumar V, et al. Trends in big data analytics [J]. Journal of parallel and distributed computing, 2014, 74(7): 2561-2573.

[7] Erevelles S, Fukawa N, Swayne L. Big Data consumer analytics and the transformation of marketing [J]. Journal of business research, 2016, 69(2): 897-904.

[8] Russom P. Big data analytics[J]. TDWI best practices report, fourth quarter, 2011, 19(4): 1-34.

[9] Arthurs J, Drakopoulou S, Gandini A. Researching youtube [J]. Convergence, 2018, 24(1): 3-15.

[10] Satish L, Yusof N. A review: big data analytics for enhanced customer experiences with crowd sourcing [J]. Procedia computer science, 2017, 116: 274-283.

[11] Li Xinrui. Research on Precise Placement and Communication Strategy of Internet Advertising [D]. Anhui University of Finance and Economics, 2021.

[12] Wen Yuhua. Research on Douyin Short Video Original Advertisement [D]. Xiangtan University, 2021.

[13] He Xuliang. Research on Influencing Factors of Bilibili Users' Behavior [D]. Wuhan University, 2019.

[14] Li Yunwei, Li Yuxin, Liang Yidan, Cheng Keni, Cai Sixiang. Selective Exposure: A Study on the Social Effects of Information Cocoons in the Background of Algorithm Recommendations——Taking the Comprehensive Interactive Video Community Bilibili as an Example [J]. China Media Technology, 2022, 9: 21-24+28.

[15] Adnan K, Akbar R. Limitations of information extraction methods and techniques for heterogeneous unstructured big data [J]. International Journal of Engineering Business Management, 2019, 11: 1847979019890771.