Journal of International & Interdisciplinary Business Research

Volume 1 *Journal of International & Interdisciplinary Business Research* | Article 4

1-1-2014

Big Data In Marketing & Retailing

Shitong Cao
University of Delaware

Ajay K. Manrai
University of Delaware

Recommended Citation

Cao, Shitong and Manrai, Ajay K. (2014) "Big Data In Marketing & Retailing," *Journal of International & Interdisciplinary Business Research*: Vol. 1, Article 4.
DOI: 10.58809/ZALA1669
Available at: https://scholars.fhsu.edu/jiibr/vol1/iss1/4

BIG DATA IN MARKETING & RETAILING

Shitong Cao, University of Delaware
Ajay K. Manrai, University of Delaware

Data is increasingly being created, stored, analyzed, and applied. Big Data is becoming an everyday phrase that appears in popular media and people's daily conversations. This paper provides a framework to define Big Data from technical and business perspectives, to present its enormous value in different fields, to share its applications in marketing and retailing, market segmentation, targeting and positioning as well in developing marketing mix. We also provide some real life industry examples, to shed light on the challenges in harnessing the potential of Big Data, and to discuss its future. Big Data will separate the winners from the losers in the business field in the future. The leading companies in the Big Data field, such as Google, Amazon, and Wal-Mart, will continue to build their competitive advantage, both in marketing and other areas, by acting on the insights developed from Big Data analysis.

Keywords: Big data, retailing, marketing mix, market segmentation, targeting and positioning, forecasting, customer life time value

INTRODUCTION

The data is anywhere and everywhere. From the geolocation information that our smart phones provide to the visited web pages that our Internet browsers record, we are only bounded by our own imagination and needs for the collection, measurement, and analysis of any type or kind of data. A recent report indicates "we create 2.5 quintillion bytes of data every day. In fact, we're creating so much data so quickly that 90 percent of the data in the world today has been created in the last two years alone" (Biehn, 2013). The explosion of the data has brought much attention and generated strong interest among businesses, academics, and general public about "Big Data". The purpose of this paper is to present a framework to provide background about Big Data, its components, applications in marketing, challenges in harnessing Big Data, and its future. This paper is organized in six sections. We begin with a brief discussion of the definition, components and application of Big Data in section II. In section III, we discuss role of Big Data in marketing and retailing. We provide applications of Big Data in market segmentation, targeting, positioning and developing marketing mix. We also discuss some specific real life industry examples in marketing and retailing. In section IV, we discuss challenges in harnessing potential of Big Data in marketing and retailing and ways to overcome the same. Next in section V, some concerns on Big Data are discussed and finally in section VI, we provide conclusions and discuss the future of Big Data.

WHAT IS BIG DATA

There is not a single interpretation of Big Data. According to the leading consulting firm McKinsey & Company, "Big Data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze." (Manyika et al. 2011) However, the business definition is more ambiguous. Typically it is referred to as the three "Vs": volume, velocity and variety. Volume is the sheer size of the dataset. Velocity is the how close the dataset is to the real time (e.g. periodic vs. real time). Variety is the different types of datasets (e.g., structured, semistructured, unstructured). A fourth V, Veracity, is also gaining recognition among the Big Data practitioners. Veracity refers to the inherent trustworthiness of the data. A more recent definition of Big Data has been put forward by Viktor Mayer-Schonerger and Kenneth Chukier in their popular book, "Big Data: A Revolution That Will Transform How We Live, Work, and Think". According to Mayer-Schonerger and Chukier, "Big Data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of values, in ways that change markets, organizations, the relationship between citizens and governments, and more." (Mayer-Schonerger and Chukier, 2013) Mayer-Schonerger and

ISSN 2332-3426 (print); ISSN 2332-3434 (online)

Chukier view Big Data as "gathering as much as possible, and if feasible getting everything: N = all". To them, Big Data is more about correlation, and less about causation.

A. Components of Big Data

1. Data Handling

Data handling is the first of the three key components of Big Data. It is related to storage and retrieval of the ever growing data and the increasing variety of data types. According to the McKinsey analysis, there has been significant growth of the data storage since 2000[3]. Also, the digital data accounted for 94% of data in 2007, compared to rest for the traditional analog data. (Manyika et al. 2011) The unstructured data has been growing much faster compared to the structured data. Specifically, the rise of unstructured data like videos, pictures, Twitter or Facebook posts, and other social media was enabled by non-relational databases that allow data to reveal its own structure. One big player in the data handling is Hadoop®, which came out of Yahoo and is an open source software framework of non-relational database that allows scaling up from a single server to thousands of machines with cheap hardware. The other major player in the processing of large quantities of data is Google's MapReduce system. Simply put, the emergence of these data handling technologies such as Hadoop® and MapReduce served as the foundation for Big Data because in the past "large datasets do not exist in any one place; they tend to be split up across multiple hard drives and computers". Also, "while traditional systems would have a delay until all updates are made, that is less practical when data is broadly distributed and the server is pounded with tens of thousands of queries per second." (Mayer-Schonerger and Chukier, 2013).

2. Data Analytics

Data analytics is the second key component of Big Data and is also commonly referred to as "data mining". Data mining is "a process of discovering and interpreting patterns in data to solve business problems" (Leventhal, 2010). There are two main types of data mining: 1. Predictive models (including multiple regression, logistic regression, and decision trees); 2. Descriptive models (including factor analysis, cluster analysis, association analysis). According to Leventha[4], There are six phases in data mining: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The McKinsey report (Manyika et al. 2011) lists some popular techniques for analyzing Big Data: "A/B testing, Association rule learning, Classification, Crowdsourcing, Data fusion and data integration, Ensemble learning, Genetic algorithms, Machine learning, Natural language processing, Neural networks, Network analysis, Optimization, Pattern recognition, Predictive modeling, Regression, Sentiment analysis, Signal processing, Spatial analysis, Statistics, Supervised learning, Simulation, Time series analysis, Unsupervised learning, and Visualization."

3. Data Visualization

Data visualization is the last key component of Big Data and is no less important than the other two key components. Data visualization refers to the ability to display and to graphically analyze massive amounts of data. In essence, it is about data communication. There are three main factors in planning successful data communication to match appropriate visualization patterns to the problems and users (Bentley, 2013): 1. The research literacy of the audience (i.e., statistics knowledge); 2. Goal orientation will dictate how to organize content on each individual view (i.e., the different things the different audiences are looking for); 3. The type of organization (i.e., Data- vs. Metrics-driven organization). For example, "metrics-driven organizations might be comfortable with an infographic, making select data very approachable. Data-driven companies want interfaces that supply greater detail and encourage exploration." (Bentley 2013) Some popular techniques for data visualization are: Tag cloud, Clustergram, History flow, Spatial information flow, Scatter plots, Heat maps, and Maps.

B. Applications of Big Data

---

[3] See Exhibit 1 in Appendices for detail.
[2] See Exhibit 2 in Appendices for detail.

https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669

Big Data has been applied in variety of sectors including public health. For example, a famous case is when Google was able to predict the spread of the 2009 winter flu in the United States, "not just nationally, but down to specific regions and even states." (Mayer-Schonerger and Chukier, 2013) Google accomplished this by comparing the 50 million most common search terms with the Centers for Disease Control and Prevention (CDC) data on the spread of seasonal flu between 2003 and 2008. After Google processed 450 million different mathematical models in order to test the search terms against actual flu cases in 2007 and 2008, Google's software found "a combination of 45 search terms that, when used together in a mathematical model, had a strong correlation between their prediction and the official figures nationwide." (Mayer-Schonerger and Chukier, 2013) The social and economic value of this capability enabled by Big Data to "predict" the spread of flu was significant because the medical resource could be more efficiently deployed and the society overall was more prepared to deal with flu season.

Besides aiding in public health, Big Data also works its magic on the streets of New York City. There are thousands of manholes in the city. In Manhattan alone, there are around 51,000. Every year, a few hundred manholes in the city start to smolder as their innards catch fire. The manhole covers sometime could explode into the air as high as a few stories and could cause serious safety incidents and business disruptions. Con Edison, the public utility company responsible for inspections and maintenance of the manholes, understood that it was not efficient to rely on chance, "hoping that a manhole scheduled for a visit might be one that was poised to blow." (Mayer-Schonerger and Chukier, 2013) The company turned to statisticians at Columbia University for help. To develop a good predictive model, the team at Columbia combed through and formatted all the available data up to mid-2008, from the messy hand-written notes of "trouble tickets", to the hundred year old records of the underground cables. The team started with 106 predictors of a major manhole disaster and was able to statistically show the two biggest factors were the age of the cables and whether the manholes had prior problems. The top 10 percent of manholes listed per the predictive model developed by the team "contained a whopping 44 percent of the manholes that ended up having severe incidents" in 2009. (Mayer-Schonerger and Chukier, 2013) Once again, this case demonstrates the value of Big Data in prediction, which enables the deployment of scarce resources more efficiently.

According to the recent study by McKinsey & Company (Manyika et al. 2011), Big Data creates values in several ways: "Creating transparency; Enabling experimentation to discover needs, expose variability, and improve performance; Segmenting populations to customize actions; Replacing/supporting human decision making with automated algorithms; Innovating new business models, products, and services". The McKinsey analysis finds that Big Data could deliver $300 billion value per year in the U.S. health care sector, and $100 billion+ revenue for service providers with "Global personal location data". [5]

BIG DATA IN MARKETING AND RETAILING

Many people mistakenly think that marketing is more art than science, and that the creative types dominate the profession. In reality, marketing is as much about science as it is about art. "A great CMO (Chief Marketing Officer) used to come with a large creative streak, or at least the ability to recognize, inspire and organize creativity. The new requirement is an ability to read and interpret data." (Kourtis, 2009). Great marketers will use both sides of their brains in taking creative risks and establishing new principles from the results of their experiments with the aid of Big Data. The rest of this section will discuss the applications of Big Data in each key area of marketing: segmentation, targeting, and positioning.

A. Market Segmentation

Marketers have long recognized the importance of segmentation to tailor their marketing mixes per the specific needs of different customer groups. Market segmentation involves deconstructing one heterogeneous market into several smaller homogenous markets in response to differing preferences due to desires for more specific satisfaction by the customers in these different homogenous markets. By using Big Data, a group of researchers developed a data mining framework for customer lifetime value (CLV)-based segmentation (Aeron, Kumar, and Moorthy, 2012). The framework uses clustering for segmentation and genetic algorithm for optimization. The methodology used for the framework mainly consists of the following[6] : 1. Deciding the appropriate CLV metric according to the context; 2. Deciding the appropriate validation metric; 3. Designing the structure of the chromosome for Genetic Algorithm; 4. Defining the fitness function for

---

[3] See Exhibit 3 in Appendices for detail.
[4] See Exhibit 4 in Appendices for detail.

the Genetic Algorithm. The dataset researchers used in the research are from a non-profit organization that uses direct mail to solicit additional contributions from past donors. By applying the data mining framework with a total of 99,200 records from October 1986 to June 1995 with each record containing 77 fields, researchers were able "to automatically segment customers (donors in this case) on the basis of CLV so that the segments are as distinct as possible. Also, the framework chooses among the set of input variables the most important variables to give distinct customer segments." (Aeron et al, 2012).

B. Targeting

Once the market is segmented to various segments, the next step is for the marketer to determine the segment(s) to target based on evaluation of segment attractiveness and company objectives and resources. Other key factors driving a target marketing strategy include degree of product variability, product life cycle stage, and competitors' marketing strategies. Big Data could play a major role in helping evaluate the segment attractiveness, such as modeling purchase behavior of customers.

Two researchers were able to develop and estimate a model of online buying using clickstream data from a web site selling cars (Sismeiro and Bucklin, 2004). The model provided superior prediction and better identified (i.e., targeting) likely buyers. To construct the model, researchers decomposed the web site into sequential nominal user tasks (NUTs).[7] The actions correspond to NUTs that site visitors must complete before purchasing online [8]. Researchers modeled the user's decision to complete each NUT as a function of variables that represent: 1. Browsing behavior (e.g., time and page views); 2. Repeat visitation to the site (e.g., return and total number of sessions); 3. Use of interactive decision aids; 4. Data input effort and information gathering and processing. Researcher evaluated three prediction models. The first model, MULT1, is the task-completion approach. The second model, SINGLE1, is a single-task model of buy behavior. The third model, SINGLE2, is the SINGLE1 model with two additional covariates, which are dummy variables that represent the completion of the first and second NUTs, respectively.

The research results indicated that "the predictive performance of all three models improves as more information becomes available about each user.[9]" "As users navigate the site and complete tasks, they provide additional information that can be used to predict purchase likelihood. The MULT1 model performs well (and much better than the single models) at the end of the sequence task, after visitors have input the personal information (i.e., after NUT2)" (Sismeiro and Bucklin 2004). In short, by applying Big Data to develop predictive modeling, researchers enabled the company to have a more targeted approach to its web site users. For example, the company could contact visitors who have exited the site without ordering if they have input personal information. The company could increase advertising exposure for those visitors who remain in intensive search mode after completing the first task. By "personalizing" the advertising content of the site, the company can augment their revenue with the sales of ad impressions and minimize any interference with the main life of the business at the same time.

Applying Big Data, another group of researchers used RFM (Recency, Frequency, Monetary) and data mining techniques with the focus on banks and other industrial partners, and analyzed the behaviors of the banks' credit cards customers (Weng et al. 2006). Research data came from credit card records from Jan 2004 to Mar 2005, which recorded what type of products or services the purchase belonged to. In total, 320 cardholders and 8,708 transaction records were obtained. The products in the data came from all types of industries including clothes, groceries, tourism, home expenses, education, and others for a total of 15 product categories. With these data, researchers did the following: 1. Analysis of the behavior of credit card customers. "Customers with different behavior are separated and so the buying features of the credit card customers can be understood, offering sales personnel a basis for planning out marketing project[10] ". 2. Analysis of the relationship between continuous buying behavior and predictive models and the choices of target customers. Researchers used two predictive models, MLE (Maximum Likelihood Estimate) and WMLE (Weighted Maximum Likelihood Estimate), to analyze the purchasing period and continuous buying behavior of credit card holders. 3. Analysis of the relationship between the active index of customers and the different predictive model: using MLE and WMLE predictive models on all types of products, researchers analyzed credit card holders' purchasing time and their active indices. WMLE considers the recency for each purchasing time into the model, while MLE does not.

[5] See Exhibit 5 in Appendices for detail.
[6] See Exhibit 6 in Appendices for the size of dataset.
[7] See Exhibit 7 in Appendices for detail
[8] See Exhibit 8 in Appendices for detail.

https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669

With these analysis, researchers developed a framework that allows "a better understanding of the buying features of the entire credit card market." Also, it enables "delineation of target customers and to predict their buying chances." The partner firms of the credit card company "will know the buying habits, likes, and buying cycle of customers so that their needs can be anticipated and satisfied, in the hope of providing the right products and services at the right time to customers who are truly in need." (Weng et al. 2006).

C. Positioning

Market positioning develops positioning for a target segment and develops marketing mix for each segment. The marketing mix is commonly referred to as the 4Ps: Product, Price, Place, and Promotion.

D. Marketing Mix

A recent study published in Journal of Marketing by a group of researchers demonstrated how Big Data could be applied to refine the 4Ps. Researchers employed text mining to extract changes in affective content and linguistic style matches (LSM) of customer book reviews on Amazon.com (Ludwig et al. 2013). Researchers gathered data using automated JavaScripts to access and parse HTML and XML pages describing books available for sale on Amazon.com. The final sample consisted of 591 books and 18,682 customer reviews. The customer review texts were automatically analyzed using the linguistic inquiry and word count (LIWC) program. "Using word counts for a given text, LIWC calculates the proportion of words that match predefined dictionaries[11]." Also, social psychology and communication research show that the manner or style in which a person communicates not only reveals personality but also elicits relational perceptions in the communication partner.

With hypothesis[12] developed per prior literature and researches, several models were constructed using the variables related to affective content and LSM. The results[13] indicate "a strong, positive, significant effect of increasing levels of positive affective content on subsequent conversion rate changes, where extreme intensity changes in affective content exhibited a quadratic (tapering-off) relationship on product conversion rates." Also, "an increasing degree of LSM in reviews related to increases in conversion rates. Finally, the interaction between the increasing degrees of LSM and positive changes in the affective content of reviews significantly predicted increase in conversion rate." (Ludwig et al. 2013)

These research results provide recommendations to online retailers to increase the effectiveness of their retail sites: "to improve conversion rates, online retailers should encourage customers to describe their product experiences in a way that reflects their emotions vividly and in a writing style that is consonant with a particular genre or product class." "Moreover, the preferred content and style alignment should be reflected in the other types of textual communication on retailer websites, such as editorial comments or product descriptions." Also, "for retailers, mining customers' particular use of affective content and function words may open up a trove of insights into their personal backgrounds, emotional states, and preferences". Another recommendation is to change the order in which the customer reviews appear. "Online retailers might order customer reviews according to their projected impact on conversion rates, such as reviews with strong affective content and an aligned linguistic style appear first."

However, researchers point out to make sure "the most visible reviews represent diverse viewpoints", because "customers discount overtly positive and/or similar reviews as biased or fake." (Ludwig et al. 2013) In short, by utilizing Big Data technique, such as text mining, researchers are able to help the organizations (online retailers in this case) to further refine its marketing mix (4Ps) to provide more timely, relevant, and targeted offerings to the customers.

E. Industry Examples

In applying Big Data to marketing in industry, a good example is Netflix (Schectman, 2012). Netflix made a successful transition from renting DVDs to providing digital video in the past few years. As Netflix prepared to offer more video via streaming, it moved its data storage from internal to Amazon's cloud. Amazon's cloud uses the highly scalable open source framework Hadoop®. This allows Netflix to expand computer resources as its needs grow. Also, "Hadoop® processing power allows the company to run massive data analyses, such as graphing traffic patterns for every type of device

[9] See Exhibit 9 in Appendices for detail.
[10] See Exhibit 10 in Appendices for detail.
[11] See Exhibit 11 in Appendices for detail.

across multiple markets". Netflix's effort to embrace Big Data improves its "reliability of video feeds on different platforms and plan for future growth of streaming movies and shows." The ability to manipulate larger data sets also enables Netflix to better analyze customer preferences, which allows it to make improved recommendations.

In another example, Toronto agency Rocket XL used Big Data and social listening to determine the types of athletes the target demo of young males were interested in and then developed a Facebook-based campaign for the Mr. Big brand (a candy bar brand) in 2012. The target audience was 17-year-old males. "Through social listening, Rocket XL was able to determine the sorts of athletes the teens hype up, entertainment they look for and kinds of things they collect. It results in a slate of creative featuring Mr. Big Deal himself, NHL start Alex Ovechkin, and a media mix that spanned online games, collectability with different packaging and quirky five-second online spots, favored by the target audience due to a shorter attention span, ultimately culminating in a specialty TV buy." (Paul, 2013) The campaign resulted in product consumption growth of 24% over a 6-month period. Also, "more than 97,000 Google search returns discussing Mr. Big and the campaign were accumulated and more than 10,000 new target-appropriate fans were engaged in Facebook".

F. Retailing

In retail, the margin is notoriously tight. However, Big Data could play a significant role in improving its productivity and operating margin. Per the McKinsey report, "in the coming years, the continued adoption and development of big data levers have the potential to further increase section-wide productivity by at least 0.5 percent a year through 2020. Among individual firms, these levels[14] could increase operating margins by more than 60 percent for those pioneers that maximize their use of big data." As the world's #1 retailer, Wal-Mart is certainly the pioneer in harnessing Big Data to serve its customers. Lately, Wal-Mart has been particularly focus on applying Big Data with its mobile strategy[15] . Per its global head of mobile division, Wal-Mart's mobile strategy is "to make mobile tools that become indispensable for our customers while shopping in our stores and online", and its goal "is to create shopping tools that become second nature to the customer, providing assistance with every part of the retail experience from pre-store planning to in-store shopping and decision making to checking out." For example, Wal-Mart's mobile app has a geofencing feature that senses when a user is in a store. It would prompt to switch to "Store Mode", which allows to scan QE codes for product price and promotions. By leveraging Big Data, Wal-Mart is "developing predictive capabilities to automatically generate a shopping list for our customers based on what they and others purchase each week." Harnessing Big Data with its mobile strategy is paying off for Wal-Mart, its "highly engaged app customers make almost four more trips a month than a nonapp users and spend 77 percent more every month". Also, "customers with a Wal-Mart app make two more shopping trips a month to our stores and spend nearly 40 percent more each month".

CHALLENGES IN HARNESSING POTENTIAL OF BIG DATA IN MARKETING AND RETAILING

According to a survey by BSkyB's subsidiary Sky IQ, which helps brands track their customer data, at least one in seven admit to feeling overwhelmed by the sheer volume of customer data available to them. "Brands can waste time gathering and sorting multiple streams of confusing data and fall into the trap of measuring outputs rather than outcomes, such as click-throughs rather than sales." Given to the enormous impact and huge potential of Big Data, it is natural to wonder what it will take for organizations to overcome the challenge in harnessing Big Data in marketing.

The McKinsey report suggests the leaders following the following steps: 1. Inventory data assets: proprietary, public, and purchased; 2. Identify potential value creation opportunities and threats; 3. Build up internal capabilities to create a data-driven organization; 4. Develop enterprise information strategy to implement technology; 5. Address data policy issues. (Manyika et al. 2011) For data analytic, the specific challenges are: 1. Deciding which data to use and where outside the organization to look. The questions about how to obtain and to integrate these data need to be answered; 2. Handling analytics and secure the right capabilities to do so. To acquire and to develop the right talent will be critical for the successful deployment of Big Data; 3. Using the insights the organizations have gained to transform the operations. It is critical that the organizations have the commitment and the right transformation model in order to reap the benefit of Big Data (McGuire, Ariker and Roggendorf, 2013). The McKinsey report states that "demand for deep analytical talent in the

[12] See Exhibit 12 in Appendices for 16 Big Data levels identified
[13] See news story by Thompson on CNBC.com, 05/22/2013.

https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669

United States could be 50 to 60 percent greater than its projected supply by 2018", and the estimated gap is 140,000 to 190,000 people.

There are usually questions about how the analytical capability should be structured within the organization in term of centralized vs. decentralized. The key is to focus on "are you providing the right service", and the structure is usually a mix between centralized and decentralized structure. The people and roles needed include business solution architect, campaign expert, data scientist, advance modelist, etc. To become data driven, the mindset and behavioral changes need to take place in the organization. To sustain the momentum for Big Data deployment in the organization, it is critical for the team to build a business case, and it needs to show business results within a reasonable timeline such as within six to nine months. It is critical to manage the life cycle of the data in term of the data architecture, its governance structure, and the data operation. Ultimately, a Big Data strategy will only be effective if the business takes ownership of it and then works in partnership with IT. Success in any Big Data strategy depends on planning.

CONCERNS ABOUT BIG DATA

A. Concerns

According to Mayer-Schonberger and Cukier, there are three main risks about Big Data: 1. Erosion of privacy; 2. Penalties based on propensities; 3. Falling victim to a dictatorship of data, where we fetishize the information, the output of our analyses, and end up misusing it.

B. Erosion of Privacy

The erosion of privacy is arguably the biggest concern of Big Data. The general public can relate to the erosion of privacy and is very personal. The concern about privacy has received a lot of attention in the media lately, especially after it was revealed that the National Security Agency (NSA) has been collecting Americans' phone call data and that the NSA is able to foil basic safeguards of privacy on the web[16] . According to the Wall Street Journal[17] , "U.S. officials have said that for all queries of the database, the NSA must show a 'reasonable articulable suspicion' that the phone number being targeted is associated with a terrorist organization." However, from 2006 to 2009, "of the 17,835 phone numbers checked against phone records, only about 1,800 were based on that reasonable-suspicion standard." Not only the public institutions but also the private organizations have raised a lot of concern about invading users' privacy. A prominent example is the discount retailer Target knowing when a woman is pregnant without the mother-to-be explicitly telling it so. Another example is Netflix releasing 100 million rental records in 2006 from nearly half a million users with the goal having an outside team improving its film recommendation system by at least 10%. Even though the personal identifiers had been removed from the data, a group of researchers from the University of Texas at Austin was able to compare the Netflix data against other public information. "A user was re-identified: a mother and a closeted lesbian in America's conservative Midwest." (Mayer-Schonberger and Cukier, 2013)

C. Penalties Based on Propensities

The power of Big Data is in its predictive capability. However, exactly due to its very nature to predict lies the second concern of Big Data: penalty based on propensity. Many law enforcement agencies are looking into Big Data to prevent crime and to spot terrorists. According to Mayer-Schonberger and Cukier, the U.S. Department of Homeland Security (DHS) has a project called FAST (Future Attribute Screening Technology) trying to identify potential terrorists by monitoring individuals' vital signs, body language, and other physiological patterns. In tests, the system was 70 percent accurate according to the DHS. This means that if the DHS is following the analysis during its daily operation, almost one out of three individuals will be mistakenly identified as terrorists. Another example is the research conducted by a professor of statistics and criminology at the University of Pennsylvania. His research method can predict whether a person released on parole will be involved in a homicide, either kill or be killed. "As inputs he uses numerous case-specific variables, including reason for incarceration and date of first offense, but also demographic data like age and gender." (Mayer-Schonberger and Cukier, 2013) Even with 75 percent probability of forecasting accuracy, it still means that should parole

[14] See news story by Perlroth and Shane on New York Times, 09/06/2013.
[15] See news story by Gorman and Barrett on the World Street Journal, 09/10/2013.

boards rely on his analysis, the boards would be wrong one out of four times. Big Data has the promise that we do what we have been doing all along, "profiling" better, less discriminatory, and more individualized. As Mayer-Schonberger and Cukier point out, "that sounds acceptable if the aim is simply to prevent unwanted actions. But it becomes very dangerous if we use big-data predictions to decide whether somebody is culpable and ought to be punished for behavior that has not yet happened."

D. Dictatorship of Data

Data is the new raw material of the century. However, we could easily enslave ourselves by blindly trying to quantify everything, without sound thinking and judgment serving as the foundation for the data. Google is the leader in the Big Data field, and runs everything according to data. Its success has been largely attributed to this data-centric strategy. However, it also trips up from time to time. For example, according to Mayer-Schonberger and Cukier, "its co-founders, Larry Page and Sergey Brin, long insisted on knowing all job candidates' SAT scores and their grade point averages when they graduated from college. In their thinking, the first number measured potential and the second measured achievement. Accomplished managers in their forties who were being recruited were hounded for the scores to their outright bafflement. The company even continued to demand the numbers long after its internal studies showed no correlation between the scores and job performance." Another example was the use, abuse, and misuse of data by the U.S. military during the Vietnam War. The Secretary of Defense, Robert McNamara, was obsessed about data and insisted on getting data on everything. Junior officers sometimes would report impressive numbers to keep their command or boost their career and make sure only to tell the higher-ups what they wanted to hear. The quality of data could be poor, biased, mis-analyzed, or used misleadingly. Sometimes, data can fail to capture what it purports to quantify. As Mayer-Schonberger and Cukier eloquently state, "Big data may lure us to commit the sin of McNamara: to become so fixated on the data, and so obsessed with the power and promise it offers, that we fail to appreciate its limitation." Recognizing the promise and importance of Big Data, we also need to appreciate people's intuition. For example, the late Apple founder Steve Job used his intuition, not data, to launch the iPod, iPhone, and iPad. "It isn't the computers' job to know what they want", he famously said, when telling a reporter that Apple did no market search before releasing the iPad.

THE FUTURE OF BIG DATA & CONCLUSIONS

A. Big Data = Big Hype?

Just as any new management theory or business practice, Big Data has its fair number of believers and doubters. A fundamental question raised is, "Is Big Data just a big hype?" There are marketing professions believe that the "Small Data" approach as opposed to "Big Data" will be more effective. "You've ended up in this cluttered mess, surrounded by information when all you really wanted was to be surrounded by answers" (Klau, 2013). According to Klau, "they (marketers) easily lose sight of the true purpose for collecting all this information in the first place: to find insights that will help them be smarter and more efficient." Fundamentally, I believe Big Data is not a big hype if the companies do it right and are able to harness Big Data to reap its reward to improve business values. Given all the challenges (e.g., lack of talents) that were discussed in previous sessions of the paper, not too many companies are doing it right at this moment as companies are just starting to recognize the potential of Big Data and to face all the challenges. Not too many companies will be able to successfully overcome these challenges to harness Big Data. More often than not, more companies will invest resource and fail. For those companies that fail to harness Big Data, they will tend to say that Big Data is a big hype. But the real reason will be due to their inability to harness Big Data, not because of Big Data itself. There will be companies that fall on the losing side of the Big Data revolution. For the companies that are already at the forefront of this Big Data revolution and have advanced their business model per benefits and insights from Big Data such as Google, Amazon and Wal-Mart, Big Data is absolutely not a big hype. To them and many others that will join them, Big Data is just as real as the air we breathe. These companies belong on the winning side of Big Data. For them, Big Data will only get even bigger each day literally.

B. The Future

Big Data is entering more and more industries. It will become integral to understanding and addressing many of our problems, not just in marketing. It has revealed areas of African slums that are vibrant communities of economic activity by analyzing the movements of cell phone users. Small Data, the world we lived in before the Big Data, will continue on and co-exist with Big Data, because not every problem or organization will have the need or capacity to use or to harness Big

https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669

Data. To address the concerns discussed previously about Big Data, the policy makers need to act. The McKinsey report states the following implications for the policy makers: 1. Building human capital for Big Data; 2. Align incentives to promote data sharing for the greater good; 3. Develop policies that balance the interests of companies wanting to create value from data and citizens wanting to protect their privacy and security; 4. Establish effective intellectual property frameworks to ensure innovation; 5. Address technology barriers and accelerate R&D in targeted areas. 6. Ensure investments in underlying information and communication technology infrastructure. (Manyika et al. 2011)

For market research, two researchers advocate the ethics of Big Data: “1. The right to be forgotten. Individuals can request that data held about them on social networking sites, which might be used for market research purposes in the future, can be deleted; 2. The right to data expiry. In addition to a general ‘right to be forgotten’, unstructured data held about individuals can be expired after a set period of time if it is of no commercial use; 3. Ownership of a social graph. Much of the value of big data is to help in the social graph where information can be collated about an individual as part of their social graph without their knowledge. Allowing individuals ownership of their social graph – information that references them and their relationships with others – prevents wider and potentially inadvertent misuse of personal data”. (Nunan and Domenico, 2013)

We may need to have a new profession that Mayer-Schonberger and Cukie advocate, called “algorithimists”. There will be both external algorithmists and internal algorithmists, just as there are external accountants and internal accounts. The responsibility of external algorithmists will be “as impartial auditors to review the accuracy or validity of big-data predictions whenever the government requires it, such as under court order or regulation.” Internal algorithimists will “work inside an organization to monitor its big-data activities.” “They oversee big-data operations, and they’re the first point of contact for anybody who feels harmed by their organization’s big-data predictions. They also vet big-data analyses for integrity and accuracy before letting them go live.” (Schonberger and Cukie, 2013)

Harnessing Big Data has allowed us to solve problems we could not previously. In marketing, Big Data has been used to develop segmentation, targeting, and positioning (4Ps). It is not the Big Hype as some people may claim. Strategically for organizations, Big Data will separate the winners from the losers. Google, Amazon and Wal-Mart and the other leading companies in the Big Data field will continue to build their competitive advantage, both in marketing and other areas, by acting on the insights developed from the Big Data analysis.

Big Data is not low hanging fruit that can be easily picked. There are challenges to harness it, from a lack of talented individuals that understand it to an organization’s illiteracy about data. The concerns about Big Data are real and personal, from erosion of privacy to dictatorship of data. However, with proper policy measurements and controls, these drawbacks will not slow down the matching of Big Data to wider and deeper applications in all different types of industries. In short, Big Data will only get even bigger!

WORKS CITED

Aeron, H., Kumar, & A.; Moorthy, J. (2012), “Data mining framework for customer lifetime value-based segmentation”, *Database Marketing & Customer Strategy Management,* Vol. 19, 1, Page 17-30.

Bentley, B. (2013), “Building User-Centric Platforms For Big Data”, *Marketing News,* Vol. 47 Issue 3, Page 46-50.

Biehn, Neil, (2013), “The Missing V’s in Big Data: Viability and Value”, *Wired*, 05/06/2013.

Gorman, S., & Barrett, D. (2013), “NSA Violated Privacy Protections, Officials Say”, *Wall Street Journal*, 09/10/2013.

Hilbert, M., & Lopez, P. (2011), “The world’s technological capacity to store, communicate, and computer information”, *Science*, Vol. 332, No. 6025, Page 60-65.

Klau, E. (2013), “Stop Hoarding Data Already”, *Adweek,* 54, 12, Page 18.

Kourtis, S. (2009), “CMOs, It’s Time for Another Renaissance”, *Brandweek,* 50, 26, Page 42.

Leventhal, B. (2010), "An introduction to data mining and other techniques for advanced analytics", *Journal of Direct, Data and Digital Marketing Practice,* Vol. 12, No. 2, Page 137-153.

Ludwig, S., & Ruyter, K. (2013), Mike Friedman, Elisabeth Bruggen, Martin Wetzels, and Geard Pfann, "More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates", *Journal of Marketing,* Vol. 77, Page 87-103.

Manyika, J., Chui, M., & Brown, B. (2011), Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers, *"Big data: The next frontier for innovation, competition, and productivity,"* Report from McKinsey Global Institute, McKinsey & Company.

Mayer-Schonberger, V., & Cukier, K. (2013), *Big Data: A Revolution That Will Transform How We Live, Work, and Think,* Houghton Mifflin Harcourt Publishing Company, New York.

McGuire, T., Ariker, M., & Roggendorf, M. (2013), *"Making data analytics work: Three key challenges"*, Interview from Insights & Publications, McKinsey & Company.

Nunan, D., & Domenico, M. (2013), "Market research and the ethics of big data", *International Journal of Market Research,* Vol. 55, No. 4, Page 2 -13.

Paul, J. (2013), "Big Data Takes Centre Ice", *Marketing*, Vol. 117 Issue 17, Page 24-28.

Perlroth, N., & Larson, J. (2013), "N.S.A. Able to Foil Basic Safeguards of Privacy on Web", *New York Times*, 09/05/2013.

Schectman, J. (2012), "Netflix Uses Big Data to Improve Streaming Video", *Wall Street Journal*, 10/26/2012.

Sismeiro, C., & Bucklin, R. (2004), "Modeling Purchase Behavior at an E-Commerce Web site: A Task-Completion Approach", *Journal of Marketing Research*, Vol. XLI, Page 306-323.

Thompson, C. (2013), "Wal-Mart Puts Its Faith in Big Data for Mobile Strategy", *CNBC*, 05/22/2013.

Weng, S., Chiu, R., Wang, B., & Su, S. (2006), "The Study and Verification of Mathematical Modeling for Customer Purchasing Behavior", *Journal of Computer Information Systems,* Vol. 47, 2, Page 46-57.

https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669

APPENDICES

Exhibit 1: Data storage grows significantly and shifting to digital

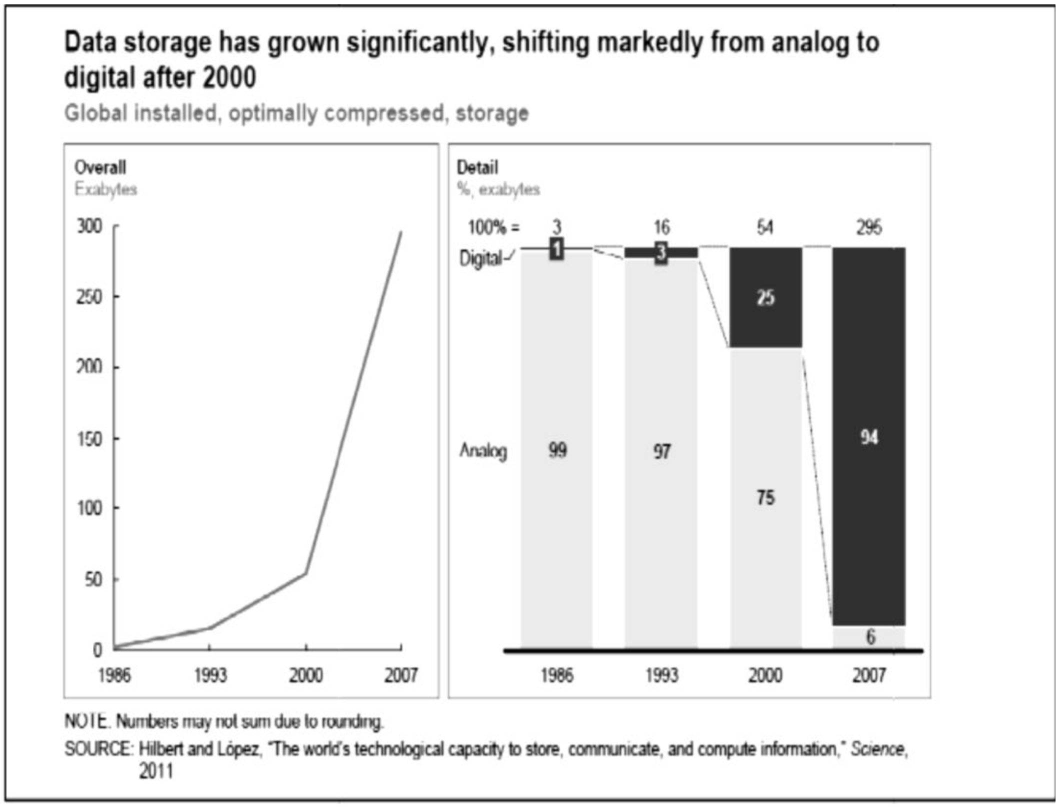


Exhibit 2: Six phases in the Data Mining process

Business understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives.

Data understanding

The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.

Data preparation

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modelling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record and attribute selection, as well as transformation and cleaning of data for modelling tools.

Modelling

In this phase, various modelling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often needed.

Evaluation

At this stage in the project, you have built a model (or models) that appears to have high quality, from a data analysis perspective. Before proceeding to the final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. A key objective is to determine whether there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

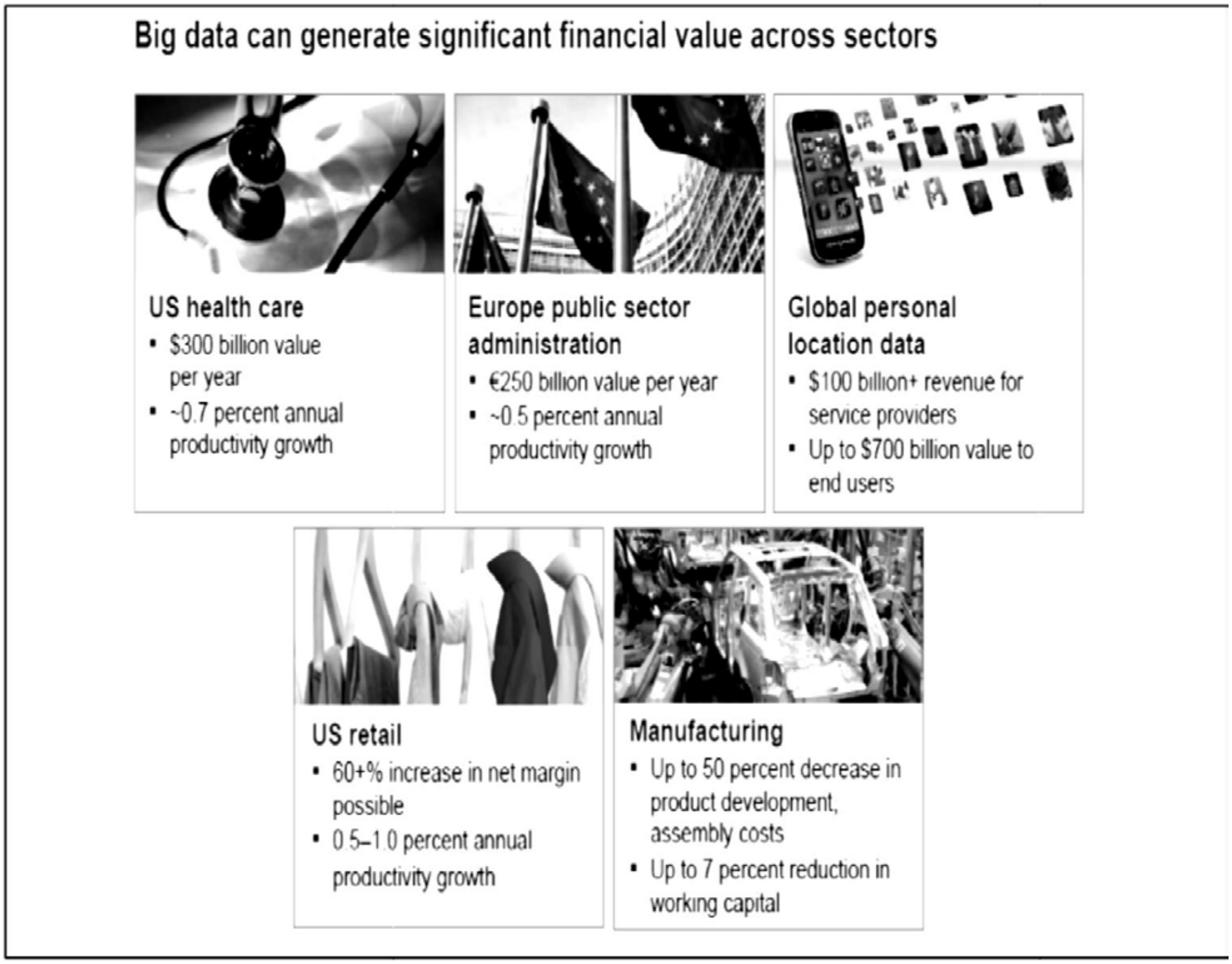
Deployment

Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the end-client can use it. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process. In many cases, it will be the end-client, not the data analyst, who will carry out the deployment steps. However, even if the analyst will carry out the deployment, it is important for the end-client to understand up front what actions will need to be carried out in order to actually make use of the created models.

Source: Leventhal, B. (2010), "An introduction to data mining and other techniques for advanced analytics", *Journal of Direct, Data and Digital Marketing Practice,* Vol. 12, No. 2, Page 137-153.

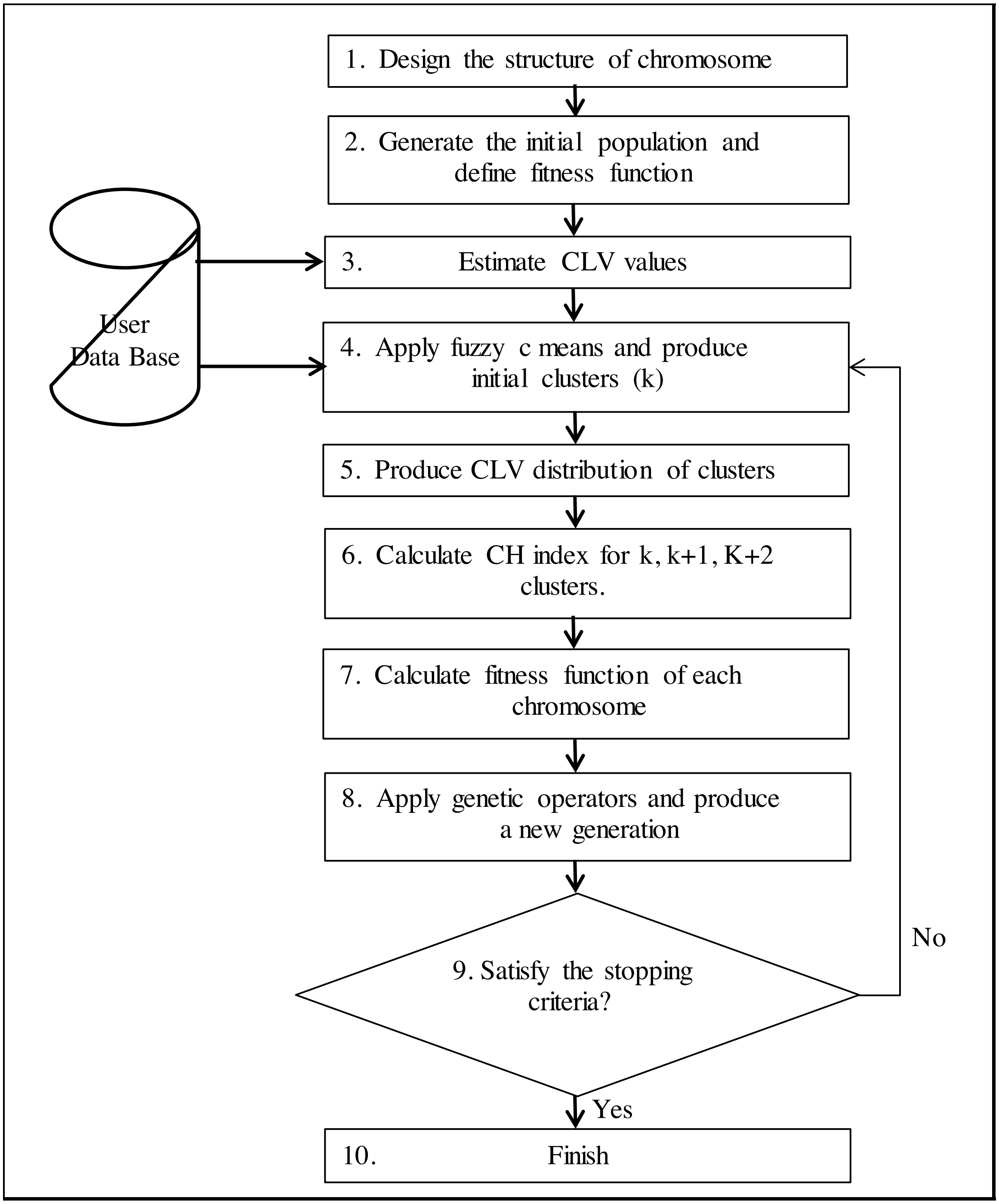
https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669

Exhibit 3: Big Data value across different sectors



Source: Manyika, J.; Chui, M.; Brown, B. (2011), Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A., *"Big data: The next frontier for innovation, competition, and productivity,"* Report from McKinsey Global Institute, McKinsey & Company.

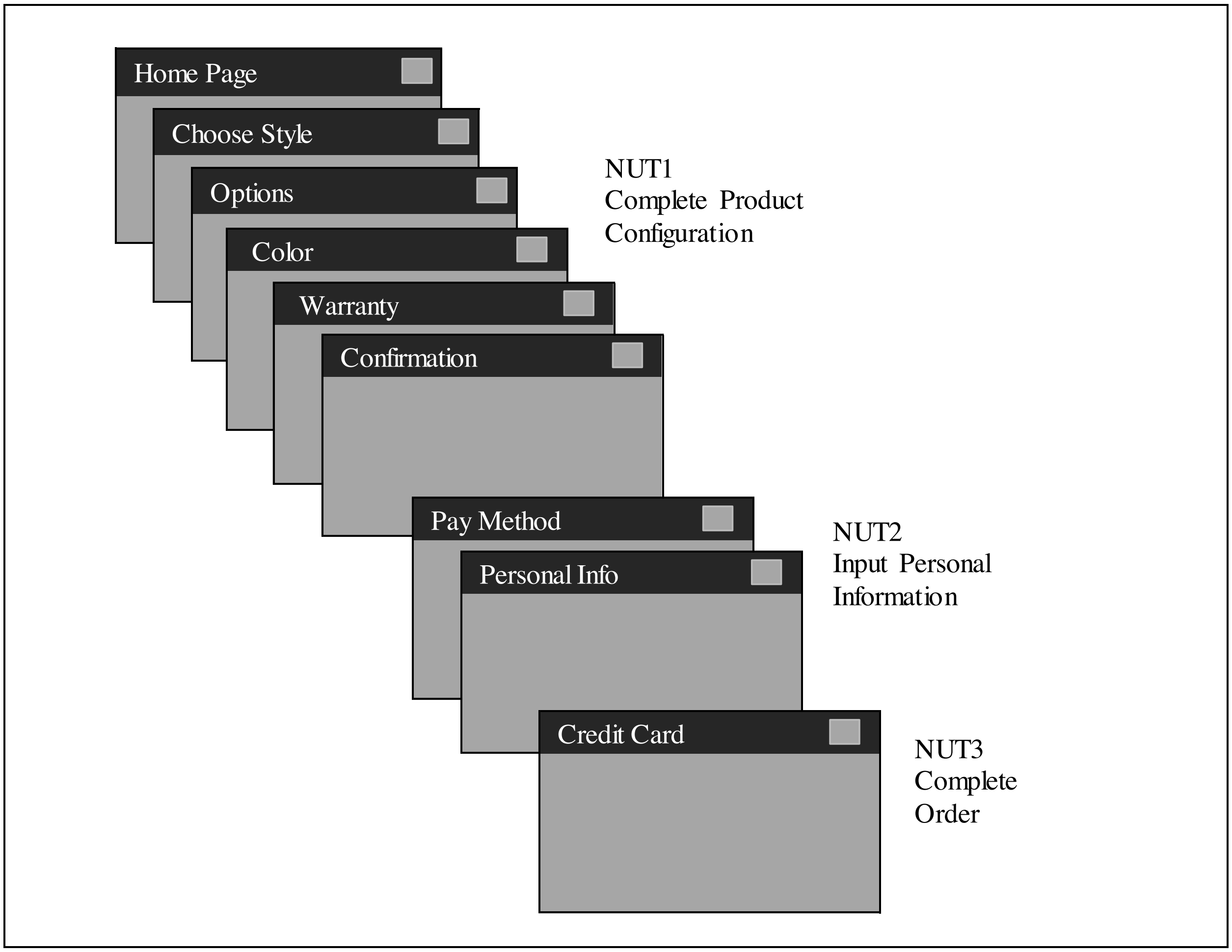
Exhibit 4: Framework for CLV-based segmentation



Source: Aeron, H., Kumar, A., & Moorthy, J. (2012), "Data mining framework for customer lifetime value-based segmentation", *Database Marketing & Customer Strategy Management,* Vol. 19, 1, Page 17-30.

https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669

Exhibit 5: The nominal user tasks (NUTs)



Source: Sismeiro, C., & Bucklin, R. (2004), "Modeling Purchase Behavior at an E-Commerce Web site: A Task-Completion Approach", *Journal of Marketing Research*, Vol. XLI, Page 306-323.

Exhibit 6: Summary statistics for NUT attrition rate

Tasks	Total Users	Users Completing Task	Percentage of Users Completing Task	Counties
Complete product configuration (NUT1)	96,498	29,238	30.30	2045
Input complete personal information (NUT2)	29,238	5753	19.68	1446
Order confirmation with provision of credit card data (NUT3)	5753	1969	34.23	689

Exhibit 7: Predictive performance

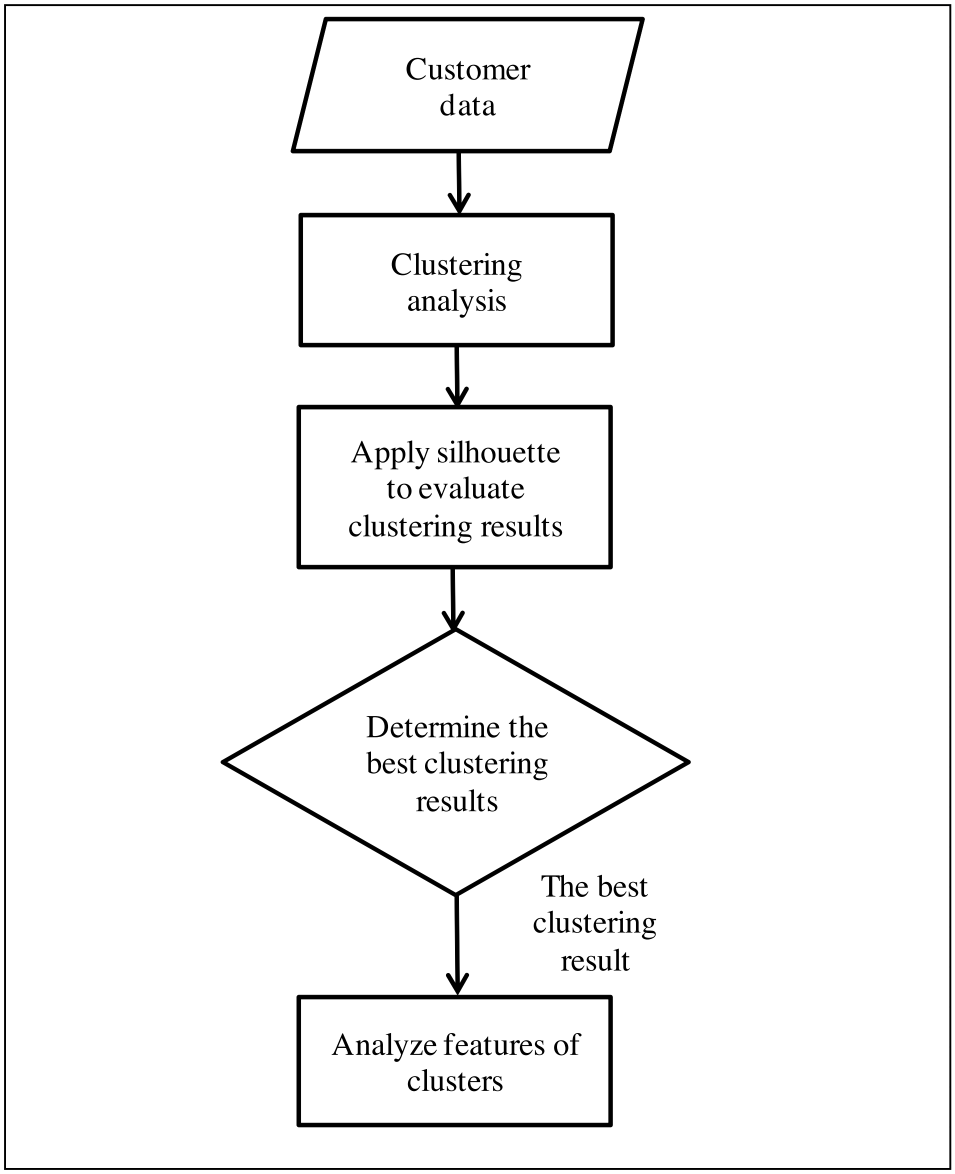
	Before NUT1			*Before NUT2*			*Before NUT3*		
	Single 1	*Single 2*	*Multi*	*Single 1*	*Single 2*	*Multi*	*Single 1*	*Single 2*	*Multi*
Overall MSE	.022	.022	.020	.065	.069	.056	.214	.209	.086
Hit Rate (%)									
Orders	.1	.1	.3	.9	.0	2.9	28.2	45.2	83.7
Nonorders	99.9	99.9	100.0	99.6	99.8	99.6	89.8	80.7	92.1
Overall	97.8	97.8	97.9	92.8	93.0	92.9	68.2	68.2	89.1

Notes: We used a .5 probability cutoff (a car ordering is predicted when the probability is at least .5).

Source: Sismeiro, C., & Bucklin, R. (2004), "Modeling Purchase Behavior at an E-Commerce Web site: A Task-Completion Approach", *Journal of Marketing Research*, Vol. XLI, Page 306-323.

https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669

Exhibit 8: Steps to analyze purchasing behavior on credit card holders



Source: Weng, S., Chiu, R., Wang, B., & Su, S. (2006), "The Study and Verification of Mathematical Modeling for Customer Purchasing Behavior", *Journal of Computer Information Systems,* Vol. 47, 2, Page 46-57.

Exhibit 9: Word categories used to calculate LSM and affective content

Category	Examples
Affective Content	
Positive affective content	*Love, nice, sweet*
Negative affective content	*Ugly, dumb, hate*
LSM	
Personal pronouns	*I, his, their*
Impersonal pronouns	*it, that, anything*
Articles	*a, an, the*
Conjunctions	*and, but, because*
Prepositions	*in, under, about*
Auxiliary verbs	*shall, be, was*
High-frequency adverbs	*very, rather, just*
Negations	*no, not, never*
Quantifiers	*much, few, lots*

Notes: We conducted the text mining using the 2007 LIWC Program (Pennebaker et al. 2007).

Exhibit 10: Hypotheses

H_{1a}: There is a quadratic relationship between changes in aggregate positive affective content in a product's reviews and changes in conversion rate. At the extremes of positive affect change, each additional increase should have a smaller impact on conversion rate change.

H_{1b}: There is quadratic relationship between changes in aggregate negative affective content in a product's reviews and changes in conversion rate. At the extremes of negative affect change, each additional increase should have a smaller impact on conversion rate change.

H_2: A positive change in LSM between a product review and the interest group's linguistic style results in positive changes in conversion rates.

H_3: There is an interaction between affective content and LSM, such that positive changs in a product review's positive affective content coupled with positive changes in LSM lead to greater positive changes in conversion rates.

Source: Ludwig, S., & Ruyter, K. (2013), Friedman, M., Bruggen, E., Wetzels, M. & Pfann, M. “More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates”, *Journal of Marketing*, Vol. 77, Page 87-103.

https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669

Exhibit 11: Results of various models on conversion rate

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Δ(Conversion rate)$_{i(t-1)}$	.917**	.712**	.725**	.727**	.763**
Δ(Conversion)$_{i(t-2)}$	−.056**	−.022*	−.021*	−.021*	−.005
Δ(LSM)$_{it}$		.004**	.004**	.004**	.006**
Δ(Affective content)$_{it}$		.051**	.050**	.050**	
Δ(LSM × affective content)$_{it}$			.002**	.002**	
Δ(Affective content²)$_{it}$				−.004**	
Δ(Positive affect)$_{it}$					.055**
Δ(Negative affect)$_{it}$					.067**
Δ(Positive affect²)$_{it}$					−.007**
Δ(Negative affect²)$_{it}$					.005
Δ(Review quantity)$_{it}$	.003*	.002*	.002*	.002*	.005**
Δ(Helpfulness)$_{i(t-1)}$	.017*	.006*	.007*	.006*	.004
Δ(Price discount)$_{i(t-1)}$	.010**	.009*	.009**	.009**	.007**
Δ(Star rating)$_{it}$	.007	−.002	−.002	−.004	−.002
Δ(Star rating variation)$_{it}$	.005*	.005*	.005**	.004*	.003*
Δ(Affective content variation)$_{it}$			.003	.002	.002
Δ(LSM × review quantity)$_{it}$			.002**	.002**	.003**
Wald's chi-square	899.93 (7)**	2128.68 (9)**	2199.56 (12)**	2234.35 (13)**	2033.87 (14)**
N	4763	4763	4763	4763	4763

*$p < .05$.
**$p < .01$.

Source: Ludwig, S., & Ruyter, K. (2013), Friedman, M., Bruggen, E., Wetzels, M. & Pfann, M. "More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates", *Journal of Marketing*, Vol. 77, Page 87-103.

Exhibit 12: Big Data levers for retailers

Function	Big data lever
Marketing	▪ Cross-selling ▪ Location based marketing ▪ In-store behavior analysis ▪ Customer micro-segmentation ▪ Sentiment analysis ▪ Enhancing the multichanel consumer experience
Merchandising	▪ Assortment optimization ▪ Pricing optimization ▪ Placement and design optimization
Operations	▪ Performance transparency ▪ Labor inputs optimization
Supply chain	▪ Inventory management ▪ Distribution and logistics optimization ▪ Informing supplier negotiations
New business models	▪ Price comparison services ▪ Web-based markets

Source: Manyika, J., Chui, M., & Brown, B. (2011), Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A., *"Big data: The next frontier for innovation, competition, and productivity,"* Report from McKinsey Global Institute, McKinsey & Company.

https://scholars.fhsu.edu/jiibr/vol1/iss1/4
DOI: 10.58809/ZALA1669