POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS

Vol. 27, No. 9, September 2018, pp. 1709–1723
ISSN 1059-1478 | EISSN 1937-5956 | 18 | 2709 | 1709

DOI 10.1111/poms.12832

Big Data and Service Operations

Maxime C. Cohen*
NYU Stern School of Business, New York City, New York 10012, USA, maxcohen@nyu.edu

This study discusses how the tremendous volume of available data collected by firms has been transforming the service industry. The focus is primarily on services in the following sectors: finance/banking, transportation and hospitality, and online platforms (e.g., subscription services, online advertising, and online dating). We report anecdotal evidence borrowed from various collaborations and discussions with executives and data analysts who work in management consulting or finance, or for technology/startup companies. Our main goals are (i) to present an overview of how big data is shaping the service industry, (ii) to describe several mechanisms used in the service industry that leverage the potential information hidden in big data, and (iii) to point out some of the pitfalls and risks incurred. On one hand, collecting and storing large amounts of data on customers and on past transactions can help firms improve the quality of their services. For example, firms can now customize their services to unprecedented levels of granularity, which enables the firms to offer targeted personalized offers (sometimes, even in real-time). On the other hand, collecting this data may allow some firms to utilize the data against their customers by charging them higher prices. Furthermore, data-driven algorithms may often be biased toward illicit discrimination. The availability of data on sensitive personal information may also attract hackers and gives rise to important cybersecurity concerns (e.g., information leakage, fraud, and identity theft).

Key words: big data; service operations
History: Accepted: December 2017 by Kalyan Singhal.

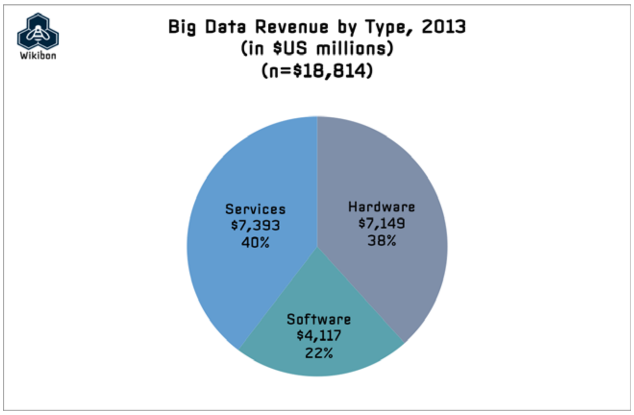
1. Introduction

This study presents a broad overview of how the service industry has been affected by the presence of big data (i.e., granular data collected and stored at unprecedented levels of variety, velocity, and volume). Today, most firms collect and store large amounts of data on their customers (e.g., personal information, demographics, social networks, geolocalization, past searches, and previous purchases), and keep a record of all their past transactions. Several experts claim that we are only at the beginning of a revolution, and that in the near future every company will base most operational decisions on data. Such a practice is often referred to as *data-driven decision making* or *data-driven algorithms*. It has become surprisingly simple to find impressive statistics on the massive size of this phenomenon. For instance, more data were created within the past 2 years than in the entire previous history of the human race. By 2020, it is predicted that 1.7 megabytes of new information will be created every second for every person on the planet. For example, Walmart handles more than 1 million customer transactions every hour. These data are imported into databases, which are estimated to contain more than 2.5 petabytes of data.[1] A recent study by Wikibon.org reported that big data will be a $50 billion business by 2017.[2] According to the same study, the big data market as measured by vendor revenue derived from sales of related hardware, software, and services reached $18.6 billion in 2013. Broken down by type, the revenue generated from big data services made it to the first place with 40% of the total market ($7.39 billion), as can be seen in Figure 1.

The vast majority of companies across different industries are aware of the potential of utilizing big data to increase their bottom line. It is now common to find a data science team in most organizations (e.g., Walmart, Staples, Zara, the New York Yankees, Marriott, American Airlines, Spotify, and Disney, to name a few). Even small startup companies often include one co-founder with expertise in data science and analytics. A 2012 survey indicated that 54% of financial services firms have appointed a chief data officer (Bean 2016). Companies are well aware of the benefits of collecting and storing past data. Interestingly, research by MGI and McKinsey's Business Technology Office revealed that firms are facing a steep shortage of talent capable of managing and analyzing big data. By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with strong analytical skills.

Most firms have invested heavily in collecting and storing data. However, it was reported that less than 0.5% of all collected data has ever been analyzed or used.[3] Thus, finding innovative and efficient ways to utilize existing data represents a major challenge for firms to unlock the benefit of big data. In the service

Figure 1 The Big Data Market as Measured by Vendor Revenue in 2013 (*source*: wikibon.org) [Color figure can be viewed at wileyonlinelibrary.com]



industry, this task is of primary importance, as historical data can help service providers learn their customers' behavior and preferences to enhance service quality and therefore, increase profit and customer satisfaction. Given the competitive landscape of the service industry, it is crucial for firms to know their existing customers (by learning their preferences from the data) in order to offer better customized services. The data on existing customers can also be used to attract similar new customers and increase the firm's market share.

In the digital era of today's economy, almost all types of services have a digital presence and are data oriented. Interestingly, firms interact with their customers via several channels. For example, users can connect and interact with a service company (e.g., an airline, hotel, or restaurant) via their Internet website, smartphone application, phone customer service, chatbot, Twitter account, etc. It is important to aggregate the multiple sources of data and to understand how a typical user utilizes the different channels offered by the firm. The challenge then is to transform the data into strategic levers that help the firm improve the current customer experience, and thus, the firm's bottom line. Big data has widely impacted management practices. According to McAfee et al. (2012): "Data-driven decisions are better decisions—it's as simple as that. Using big data enables managers to decide on the basis of evidence rather than intuition. For that reason it has the potential to revolutionize management." A similar claim also applies to service operations, which should strive to unlock the powerful information hidden in big data.

In this study, we discuss how the presence of big data has transformed services. In particular, section 2 focuses on financial services (credit cards, online payments, peer-to-peer lending, and trading services). Section 3 discusses the transportation and hospitality sectors, with an emphasis on ride-hailing and marketplaces, as well as traditional firms, such as airlines and hotels. Section 4 considers services delivered by online platforms. In section 5, we present some of the main pitfalls and risks due to the abundance of big data. Finally, our conclusions are reported in section 6.

For each topic and industry, we convey some of the recent changes induced by the presence of large datasets. We discuss in greater detail companies with which the author collaborated through research, consulting, and/or informal discussions. The ideas presented in this study are by no means exhaustive and portray only a partial representation of the actual situation which is evolving every day. Our goal is to make the reader aware of the impact of big data on services, as well as to discuss some of the marketing and operational tools commonly used to improve the decision-making process by exploiting historical data. The focus of this study is less on the technical aspects and more on the high-level picture. Before presenting concrete examples from the aforementioned sectors, we briefly discuss several tools and mechanisms the service industry uses to analyze and exploit big data.

1.1. Tools and Mechanisms

Companies seek to use their data to identify trends, detect patterns, and know better their customers' habits and preferences. As big data holds the potential to describe customers with a high accuracy, many firms use their data to refine their definition of the Ideal Customer Profile (ICP). Furthermore, one can use past digital information to define target audiences (age, location, education, income, interests, etc.). By observing users' habits and online search results, the firm can identify the type of customer who is attracted by certain products. Exploiting big data provides insights into how to lower the Customer Acquisition Cost (CAC), increase the Customer Lifetime Value (CLTV), reduce customer churn, and manage several other customer-driven metrics.

The process of collecting, storing, and analyzing data is based on several steps, and each step can be performed, using specific tools and software. These steps include data collection, data storage and management (e.g., Hadoop and Cloudera), data cleaning (e.g., OpenRefine and DataCleaner), data mining and analysis (e.g., IBM SPSS Modeler and BigML), data visualization (e.g., Tableau), and data integration (e.g., Talend). In this study, we focus on the data mining and analysis step. Next, we describe mechanisms (or methods) commonly used in the service industry that leverage the potential information hidden in big data.

- **Real-time personalization:** Companies aim to send the right offer to the right user at the right time. Big data can be used to effectively personalize offers, prices, displayed assortments, and the order of online searches. Personalization can be performed dynamically, as well as at the user level. Depending on the context, personalization can utilize customers' attributes, such as geo-localization, demographic information, time of the day, etc. Today, differentiating pricing strategies at the customer-product-time level and optimizing prices using big data have become ubiquitous. Several academic papers were written on this topic (see, e.g., Amatriain 2013, Golrezaei et al. 2014).

- **Targeted promotions and campaigns:** Sending promotions to existing or new customers can be expensive and often results in low conversion rates. Using large historical data sets can help guide the design of future promotion campaigns. For example, firms need to decide which types of customers to target, and what are the most important features (e.g., geo-localization, demographics, or past behavior). Search engine optimization and targeted marketing campaigns (email and mobile offers) are two domains where big data is having a significant impact. Designing targeted promotions has been extensively studied in the academic literature (see, e.g., Andrews et al. 2015, Arora et al. 2008, Fong et al. 2015).
- **Pattern and trend identification:** This process, sometimes also called lead scoring, involves the use of historical data to identify existing trends and predict future outcomes (see, e.g., Thomas et al. 2017 and the references therein). Companies try to identify shopping trends (e.g., by using Google Trends), so they can highlight their top-selling products and eliminate the underperforming ones (Choi and Varian 2012). Other examples include identifying demographic trends, which allows businesses to better cater to specific groups. Advanced statistical and machine learning algorithms are routinely used to search for new patterns and correlations in the data. Subsequently, such patterns are used to inform future operational decisions. One common technique is *association rule learning*, which aims to discover interesting hidden relationships among several attributes in large databases. One application is market basket analysis, in which a retailer can determine which products are frequently purchased together and then use this information for marketing purposes.
- **Customer segmentation:** This consists of clustering customer data to create customized marketing campaigns that cater to several specific groups of users. Examples include recommendation engines that create value for customers by reducing their search and evaluation costs (e.g., Amazon and Netflix). Firms try to segment users based on several attributes in order to leverage the data on existing users to make decisions for similar new customers (see, e.g., Bobadilla et al. 2013, Zhou et al. 2010).
- **Predictive modeling and analytics:** Predictive analytics is a commonly used statistical technique to algorithmically predict future outcomes based on historical, current, and even social data (Cui et al. 2017). In practice, predictive analytics can be applied to a wide spectrum of disciplines–from predicting the failure of an aircraft engine based on the data stream from sensors to predicting customers' next actions based on their previous purchases and their social media engagement. One popular application is fraud analytics, where the data are used to detect fraudulent activities.
- **Text, audio, and video analytics:** Text analytics (or text mining) refers to techniques that extract information from textual data, obtained from online reviews/forums, emails, survey responses, social network feeds, and call center logs. Text analytics involves statistical analysis, computational linguistics, and machine learning. Audio or speech analytics analyzes and extracts information from unstructured audio data. Currently, call centers are one of the primary application areas of audio analytics. Recently, video analytics has also received a lot of attention.

In the next three sections, we report several concrete examples of services from different sectors that use at least one of the mechanisms above.

2. Financial Services

In this section, we discuss how big data has been shaping some of the financial services. We first comment on the impact of credit cards applications, which is a multi-billion dollar industry. Second, we discuss the impact on online payment platforms that have emerged in recent years. Third, we consider the market of matching borrowers with lenders (peer-to-peer lending), which is a growing industry powered by large datasets. Fourth, we briefly discuss the impact of big data on trading and investment services. More details on operations in financial services can be found in Xu et al. (2017) and Pinedo and Xu (2017).

2.1. Credit Cards

Banks and financial institutions are constantly seeking to improve their credit card departments. Credit card services yield a comfortable margin and have become a very lucrative business. For example, in 2014, American Express made a net interest income of approximately \$5.8 billion.[4] From the bank's perspective, credit cards generate revenue by charging a membership fee, a small fee per transaction to merchants (called the interchange fee, which typically ranges between 1% and 6% of the transaction amount), international fees on transactions in a foreign currency, and exchange rate loading fees. In many cases, banks also charge a high penalty fee for late or overdue payments, a fee for exceeding the credit limit (called the over-limit fee), as well as relatively high interest rates on outstanding balances. From the user's perspective, having a credit card is almost a necessity (at least, in the United States). The variety of credit card options has exploded in the last two decades. It is now common for companies to partner with a bank in order to offer their own credit card (e.g., Amazon, American Airlines, Macy's, Costco, Best Buy, Toys "R" Us, and the recent example of the Uber VISA card). The number of credit cards in circulation and the number of active users have also exploded. In 2015, the number of credit cards from the four primary credit card networks (VISA, MasterCard, American Express, and Discover) was 636 million.[5] In addition, Americans made 33.8 billion credit card transactions in 2015, for a total value of \$3.16 trillion.[6] As of November 2016, a typical American held an average of 2.35 credit cards.[7]

Picking the right credit card (either for personal or business purposes) can save users thousands of dollars each year on interest, helps book travel for free, as well as earns significant rewards (e.g., receiving cash back, collecting airfare miles, and accessing airport lounges). It is also important for users to use their credit cards wisely, as it affects their credit score, which is crucial for loan applications. With the unprecedented volumes of data on users and past transactions, credit card companies seek to send targeted offers to users to increase their profits. In particular, companies use data-driven models to predict the default probability of users, as well as their propensity to spend. Such companies often use their data in order to guide decisions, such as (i) which individuals to offer a credit card, (ii) the credit limit, (iii) when and by how much to increase the credit limit for existing users, and (iv) which benefits or rewards to offer in order to increase the conversion rate.

Consider the concrete problem faced by a credit card company that needs to decide whether to issue credit cards to a specific set of users. The first goal is to assess their risk of default. Note that this task can be challenging as the historical data is available only for individuals who were given a credit card in the past. Consequently, for users who were never issued a credit card, it is not possible to know if they would have defaulted. If the population of people applying is "similar" to the population who were issued cards in the past, one can use the past information to infer future decisions. But very often, this is not quite the case. In particular, there are a large number of applications from customers who are completely unbanked, with no credit history (e.g., new immigrants and students who get their first job). The presence of large datasets can help identify similar users (by using clustering techniques) and leverage some detailed information on similar users.

An additional business decision related to credit cards is to decide whether to give a credit limit increase to existing customers. Existing users may request a credit limit increase every several months. Then, the credit card company needs to decide whether or not to approve such a request. On one hand, the company wants to increase the spending power of users, as it may potentially enhance the firm's profits. On the other hand, this can lead to a higher risk of defaulting, and such decisions are subject to strict laws and regulations. More precisely, credit card companies cannot grant infinite credit limit increases as the companies must keep a portion of their capital on hand to cover the credit they are issuing. Consequently, with the credit limit constraint, the firm wants to increase credit limits for individuals who are the most likely to spend, while minimizing the default risk. Solving this optimization problem calibrated with historical data is definitely not a simple task.

In addition to credit card companies, banks face similar data-driven decision-making problems. Consider, for example, a bank that needs to decide which financial product to suggest next to a particular client. Banks always try to persuade customers to acquire new products (e.g., a credit card, a savings account, a mortgage, or a private wealth account). However, advertising a new product can often be expensive. Therefore, several banks put great efforts into carefully selecting whom to advertise to, and which product(s). Similarly, banks need to decide which price to offer for long-term savings accounts. Many individuals have at least one savings account. When the term ends (e.g., every year) and the account rolls over, the user has to decide whether to renew, and this decision depends on the rate offered by the bank. Today, banks often rely on past data to solve this problem. Several related works can be found, see, e.g., Bharath et al. (2009).

Interestingly, some credit card companies sell (anonymous and aggregated) customers' data to other businesses, such as retailers that would like to garner better insights into consumer spending habits. The data can be aggregated by ZIP code, which informs retailers what areas are more likely to make purchases. Alternatively, credit card companies can sell data to advertisers that can use this information to target specific users with ads. Similarly, companies, such as creditkarma, sell aggregate information to credit card companies and to advertisers. Creditkarma is a personal financial management platform that offers free credit scores (with weekly monitoring), tax preparation, tools for identifying and disputing credit report errors, and tools for simulating credit score evolution. The revenue stream of such (free to consumers) platforms is typically covered by targeted advertisements for financial products (e.g., credit card recommendations based on users' credit profiles).

2.2. Online Payments

Online payment systems can be based either on fiat currencies or on virtual currencies. Traditionally, small payments in the United States, as well as in many other countries, were made through checks. Launched in December 1998, PayPal quickly became the leader in accepting and receiving online payments. It appears to be the de facto online payment solution for online customers, freelancers, and small business owners. For example, many transactions on eBay are performed via PayPal. Similarly, a large number of websites that aim to request donations use PayPal. PayPal's shares increased by nearly 80% between January and October 2017 with $85.8 billion in market capital (October 26, 2017). In October 2017, PayPal launched a new product called PayPal for Marketplaces. This new modular system is designed for businesses that operate online marketplaces (e.g., ride-sharing and room rental platforms, crowdfunding portals, peer-to-peer e-commerce sites, and online gaming). In recent years, technological advances have opened the door for a number of competitors to challenge PayPal by offering cheaper fees, faster transactions, and enhanced security. Over the last several years, new payment systems proliferated. One can count more than a dozen of alternatives. Examples include Stripe, Due, Apple Pay, Google Wallet, Payoneer, Square, Alipay, Amazon Pay, Skrill, WePay, and Venmo; a more exhaustive list can be found online.[8] Such systems may allow users to complete payments via email or by using their smartphones. Consider, for example, the service offered by Venmo (which was acquired by Braintree in 2012, which was itself acquired by PayPal in 2013). Venmo is designed for quick and small payments (after verifying their identity, users can send up to $2999.99 during each seven-day period). Although the number of users is not publicly reported, the dollar amount in transactions is quite impressive. Venmo handled $17.6 billion in transactions in 2016, $6.8 billion in transactions in Q1 of 2017,[9] and more than $8.0 billion in transactions in Q2 of 2017.[10] The main competitive edge of Venmo lies in its social dimension. In particular, a popular use case is when friends conveniently split bills (e.g., for meals, movies, rent, or trips). Venmo allows users to log in to the platform using Facebook, thus providing access to social network data to the provider. When a user completes a payment transaction, the transaction details (the payer, receiver, amount, and specifics of the expense) are shared on the user's "feed," so that other friends can also see it.[11] In addition, Venmo encourages social interactions on the platform through likes and comments on each transaction. Consequently, the richness of the data available to a platform like Venmo is striking. The platform has access to the network of friends, to the types of relationships and mutual interests people have (e.g., going to watch a movie on a weekend), and to all the pairwise money transactions. Monetizing this data is a challenge but if done properly, can lead to very high profits.

2.3. Peer-to-Peer Lending

In today's economy, borrowing and lending often occur online, especially when the borrowing party does not have a high enough credit score. Borrowing and lending (peer-to-peer) can take place between two individuals without involving a banking institution. Peer-to-peer lending refers to the practice of lending money to individuals or businesses through online services that match lenders with borrowers. The borrower must be able to provide sufficient information about his or her creditworthiness on the online platform in order for the lender to be able to assess the credit risk. The interest rates can be set by lenders who compete for the lowest rate on the reverse auction model or fixed by the intermediary firm based of an analysis on the borrower's credit.

For many years, private individuals have already been offered the option to secure mortgages on their homes through websites such as LendingTree.com. This platform is an online lending exchange that matches individuals with given credit rating scores (FICO) with established banking institutions that compete for business. Such transactions are made possible thanks to reliable credit scores provided by three credit rating agencies (Equifax, Experian, and TransUnion) that collect extensive financial information on individuals. At times, the borrower may be a small company (e.g., a startup) that would have difficulties to obtain financing from a banking institution. Such a company may also resort to crowdfunding to obtain financing in the form of a loan or an equity stake in the company.

LendingClub is the world's largest peer-to-peer lending platform.[12] More precisely, it is an online lending platform that offers loan trading on a secondary market and enables borrowers to create unsecured personal loans between \$1000 and \$40,000. The company claims that \$28.75 billion in loans have been originated through its platform up to June 30, 2017.[13] Each loan displayed on the website includes information about the borrower (e.g., FICO score, credit history, debt-to-income ratio, and personal income), the amount of the loan, the loan's grade, and the loan's purpose. Investors earn money from the interest, whereas the platform charges borrowers an origination fee and investors a service fee. The amount of historical data available to the platform is enormous. This data (which is partially made publicly available) is transforming the lending industry and has incentivized several investment firms to enter this market.

2.4. Investment and Trading Services

Investment and trading services, which often occur through online platforms, are other prominent financial services that have been affected by big data. Users can easily create accounts allowing them to trade securities, stocks, bonds, and exchange-traded funds (ETFs). Such platforms have access to unique datasets. For example, the platform can have access to how often users connect to the platform and which types of financial products users monitor. Subsequently, the platform can sell such aggregate information to advertisers. In addition, the platform can send targeted offers to its users, such as free webinars, referral promotions, and online workshops on different topics. An interesting example is the platform eToro, which is a social trading and multi-asset brokerage company.[14] Users can trade currencies, commodities, indices, and contract for difference (CFD) stocks online. The unique characteristic of eToro's platform is that users can decide to follow investors by replicating the same portfolio investment. The slogan on their website reads as follows: "Join the Social Trading revolution! Connect with other traders, discuss trading strategies, and use our patented CopyTrader technology to automatically copy their trading portfolio performance." Their data is very rich as it includes performance data and risk measures for each user at each point in time. In addition, eToro has access to fascinating data on social connections and influences among the different users. Each user is rated with a risk factor, the gain or loss in percentage during the last 12 months, and the variation in gain or loss during the last trading day. One can also access historical statistics on the performance and the current portfolio (i.e., open trades). Finding effective ways to monetize such data is an interesting challenge, and several companies are working on this type of problem.

3. Transportation and Hospitality

In this section, we discuss some of the recent disruptions in the transportation and hospitality industries that were partially driven by the presence of big data. The world is clearly moving toward personalization in these sectors, implemented through a large-scale decentralized system. Most service providers aim to constantly improve their customer service by collecting large relevant datasets on users. We first consider transportation services (with a focus on ride-hailing platforms), and then consider hospitality services (with a discussion on hotels and online marketplaces, such as Airbnb).

3.1. Transportation

On-demand ride-hailing platforms have changed the way people commute and travel for short distances. Several well-known players in this market are Uber, Lyft, Didi Chuxing, Grab, Ola, Via, Gett, and Juno, to name a few. In October 2016, it was reported that Uber had 40 million monthly riders worldwide.[15] Today, using this type of transportation service has become the norm in most major cities (e.g., Uber now operates in more than 600 cities around the world). During the first few years, growth was moderate, but within the last 2 years, this industry has expanded rapidly. For example, it took Uber 6 years to complete their first billion rides (from 2009 to 2015) but only an additional 6 months to reach their two-billionth ride.[16] This means that during the first 6 months of 2016, the company was providing an average of 5.5 million rides a day (or 230,000 an hour). This type of statistics illustrates the impressive scale and growth of the ride-hailing industry. More importantly, ride-hailing platforms collect a massive amount of granular data at a very large scale. Each transaction (in this case, ride request) comes with a stream of information: rider ID, drop-off/pick-up times and locations, number of passengers, day of the week, price offered, weather conditions, waiting time, type of car requested, and much more. For example, services like Uber allow riders to split the fare with friends. Thus, this provides information on users' social networks. Platforms can collect and store information on geo-localization, willingness to pay, willingness to wait, as well as many other features related to their customers. Using this rich data on every single user remains challenging but has unprecedented potential in terms of increasing the service personalization and the long-term profits. For example, if the platform knows that some users do not mind waiting for rides, whereas others do, this information could be potentially used in the "dispatch algorithm" (i.e., deciding in real-time which drivers are assigned to which ride requests). These platforms always try to find new ways to exploit the data in order to improve service, retention, and revenue. A recent example is related to geo-localization. When a user requests a ride and does not accept it (e.g., the price was too high), the application can notify the user a few minutes later that the price quote is now lower (while knowing the rider's exact position). The firm can also use the data to decide how to send targeted promotions to its users. By leveraging available fine-grained data, promotion and referral campaigns can now be customized to a very high degree of granularity. Since such platforms often operate in an on-demand supply mode, they also collect a vast amount of data on workers/drivers (vehicle type, sometimes demographics information, geo-localization, work hours, sensitivity to promotions, etc.). As a result, the platforms can refine their algorithms and increase the efficiency of their operations by using this data to create better incentives for both riders and drivers. This topic is a very active research area in the operations management community (see, e.g., Bimpikis et al. 2016, Chen and Hu 2016, Chen and Sheldon 2016, Cohen and Zhang 2017, Hu and Zhou 2017, Tang et al. 2017, Taylor 2017).

Recently, taxi services have dedicated great efforts to modernize their operations to better fit into today's economy. For example, in several cities, taxi rides can now be directly ordered from a smartphone application, and the payment (including the tip) can be completed either via the application or in person. One such company based in the United States is Curb.[17] On their website, one can read: "Curb is the #1 taxi app in the United States that connects you to fast, convenient and safe rides in 65 cities (50,000 Cabs—100,000 Drivers)." Other similar examples include Ace Taxi in Cleveland, Ohio, Talixo in Germany, and taxi.eu which operates in 100 European cities. These companies offer taxi rides by using an online platform, and therefore, can easily collect data on previous transactions. Historical data can help taxi companies improve operational decisions, such as dispatching drivers across the city, predicting demand in real-time, and sending targeted offers. Most optimization and prediction algorithms used by such platforms are data-driven and are tuned very often in order to dynamically capture the high-paced changes observed in the data.

Interestingly, several platforms go beyond just passively collecting data. In particular, several firms design and run a multitude of micro-experiments with the goal of generating high-quality data. A typical platform can decide to run a series of carefully designed experiments (often called *A/B* tests), in order to validate intuition and gain important novel knowledge on users' behavior. For example, are users more likely to react to promotions sent in the morning or in the evening? To answer such a question, the firm can design a small experiment and randomly send promotions to two samples of users. Then, by testing the statistical significance of the results, the platform can gain important knowledge that will be valuable moving forward. Today, Microsoft and several other leading companies, including Amazon, Booking.com, Facebook, and Google, each conduct more than 10,000 online controlled experiments annually, with many tests engaging millions of users (Kohavi and Thomke 2017). Startups companies and firms without digital roots, such as Walmart, Hertz and Singapore Airlines, also run this type of test regularly, albeit at a smaller scale. For more details on this topic, we refer the reader to the recent article by Kohavi and Thomke (2017) and to the paper by Kohavi et al. (2013).

Airline companies have also been greatly affected by big data. The pricing and scheduling decisions for flights are often controlled by data-driven algorithms. Airline companies collect rich datasets on customer transactions (customers can now easily be identified via frequent flyer numbers). Firms subsequently use this data to customize price offerings and to enhance customer loyalty. The demand prediction for each leg is also performed by using large datasets which include previous performance, weather conditions, as well as many additional factors. Today, airlines make the majority of their operational decisions (scheduling, pricing, inventory, staffing, etc.) based (at least partially) on historical data. From the customer perspective, things have also evolved significantly. The increased level of competition and the explosion of reservation systems (e.g., Kayak, Orbitz, and Expedia) allow consumers to easily compare the different alternatives. Some of these reservation systems even offer advice on whether to book now or to wait for a potential price decrease. These reservation systems have access to very fine-grained data about users: geographical location, IP address, browser used, mobile or desktop platform, past searches, previous reservations, number of clicks, and so on. The systems can sell aggregate information to advertisers and use some of this information to discriminate searches and prices among users (see more details on this topic in section 5.3).

Big data have also affected the transportation industry from a completely different angle. Manufacturers and in particular, large aeronautics companies, such as Boeing and Airbus, now routinely use data obtained from sensors to manage the maintenance of their aircrafts, and to create after-sales personalized services (e.g., proactive maintenance and special monitoring). More precisely, they place hundreds (if not thousands) of different sensors to collect information in a very fine-grained fashion. Those sensors are very sophisticated and can often record measurements as fast as every second. They are located in different parts of the aircraft and typically measure the temperature, the humidity level, the vibrations, as well as various physical and mechanical properties. The data collected from these sensors gives rise to very large time series (such datasets are often hard to store and to visualize). Subsequently, the firm's goal is to analyze these datasets in order to improve the current maintenance strategy. The ultimate goal is to send specific aircraft parts for maintenance before a critical issue occurs, but at the same time not too early. This practice involves very high costs and risks, and thus, the potential impact of properly analyzing such data is tremendous. Firms that offer this type of service to airlines have a unique competitive advantage. A very similar story (albeit, at a smaller scale) is present in the automobile industry, which has also started to use a large number of data sensors.

3.2. Hospitality

Today, hotels collect and store as much information as possible on their customers. Nevertheless, it can be challenging to efficiently exploit the data as individuals often use different email addresses (e.g., business versus personal). One common technique to ease the data collection process is to offer a loyalty card to customers. Very often, the loyalty program can be shared among several hotel partners (e.g., the Starwood Preferred Guest network includes Sheraton, Le Meridien, Westin, W Hotels, and several other hotel chains). This allows the firm to identify each user by a unique identifier. After each stay, hotels can record the special requests and preferences (e.g., vegetarian or booking room dining services). During the next stay, the hotel can then accommodate the customer's needs in a more effective and personalized fashion.

Most hotels also use historical data to design and send targeted promotional offers. For example, hotels can collect data on which types of customers are likely to accept which type of promotional offer, and at what time of the year. Then, by leveraging the data from past offers, the hotels can decide the specifics of the next campaign. For example, hotels in the Intercontinental Hotels Group use price optimization tools (see Koushik et al. 2012). Another example is the Carlson Rezidor Hotel Group that uses data-driven methods to maximize revenue (see Pekgün et al. 2013). This type of practice is particularly relevant to hotels in Las Vegas (or other casino resorts), which use data-driven predictive algorithms to infer the spending capital of each potential customer. Such practices are often taken very seriously as they can drive a relatively large portion of the hotel's revenue. This topic has been extensively studied in the academic literature (see, e.g., Kilby et al. 2005, Lucas and Brewer 2001). In addition, a large number of patents were issued on this topic during the last two decades. Having access to larger datasets can only make this lucrative practice more exciting. It is now common for hotel groups to hire a team of analysts who are dedicated to improving their data-driven algorithms. In addition, with the explosion of online travel search websites (e.g., Kayak, Orbitz, and Expedia), hotels need to decide at each point in time the portion of reservations to assign to those channels, as well as the pricing policy. Managing operational decisions for cruises has also been affected by the presence of large datasets (see, e.g., Gibson 2006). Today, several cruise companies have a department fully dedicated to developing and improving the management and use of their data.

It does not seem reasonable to end this section without mentioning online marketplaces such as Airbnb, Homeaway, Homestay, VRBO, Vacasa, and Flipkey, to name a few. These online platforms allow people to lease or rent short-term lodging, including vacation and apartment rentals. These companies typically do not own any lodging but receive a percentage service fee (commission) for every booking from guests and hosts. As of October 2017, Airbnb has more than 3 million listings in 65,000 cities and 191 countries.[18] The price is typically set by the host and can depend on the time of the year, the day of the week, the number of similar listings, the amenities available, the number of nights, etc. The amount of data gathered by such platforms is impressive, as they can collect data on both the users and the properties. This data allows the platform to deliver a better service by recommending prices to the host, and improving the ranking of the different options for each browsing user (based on previous preferences). It also leads to more transparency for the industry (users can write reviews of good and bad experiences so that other users' knowledge increases considerably before booking). Such platforms often hire senior data analysts who are constantly working on exploiting historical data to improve future tactical and operational decisions.

4. Online Platforms

In this section, we focus on services which are offered via online platforms. These services are also greatly affected by the presence of big data. Today, many firms interact with their customers via online platforms. This is true for transportation services (ride-hailing), as discussed in section 3.1. Other examples include health and medical services (e.g., Zocdoc), dating services (e.g., match.com), recruiting services (e.g., CornerJob), restaurant reservations (e.g., OpenTable), food delivery services (e.g., Grubhub and

Slice), delivery and home services (e.g., TaskRabbit and Handy), and subscription services (e.g., Stitch Fix and Blue Apron). Services delivered by online platforms have transformed a big part of the service industry. These companies can collect vast amounts of fine-grained data on customers' habits and preferences with the goal of improving service customization. The ultimate objective is to offer the optimal product for each customer, and vary the prices dynamically to maximize long-term revenue and retention. Learning user preferences (for food, clothing style, dating affinities, etc.) can be performed by analyzing survey data, questionnaires, online reviews, data from social networks (friends, pictures, and interests) and matching or clustering users with other carefully selected similar users. This can be accomplished only by leveraging and analyzing the large amounts of historical data. For example, to use the service offered by Stitch Fix, customers fill out a survey about their style, pay a $20 up-front styling fee, and then receive five clothing items tailored to their taste. Stitch Fix runs relatively large data science operations that leverage the data from in-depth surveys to increase the accuracy in styling choices for customers. This type of firm owns a valuable growing dataset of detailed customer preferences and product trends and is constantly working on refining its data-driven algorithms. As mentioned before, such datasets can be sold to online advertisers. Online services such as music (e.g., Spotify) and movies (e.g., Netflix) also dedicate significant efforts to collecting and exploiting large amounts of data. One of the basic challenges is to accurately learn users' preferences from past usage in order to provide effective recommendations. These companies are actively hiring top data scientists to develop efficient algorithms that operate in real-time at a very large scale.

Apart from these relatively new examples, one can find similar data-driven practices by companies like Amazon, Facebook, YouTube and many others. Online retailers (e.g., Amazon) know when a user recently visited their website and browsed a specific product. Then, such a user may have a high valuation for being shown an ad for this specific product. This common practice is known as remarketing or retargeting, and can generate significant revenue for retailers and for advertisers. Facebook has access to endless data on its users, and strives to exploit this data to optimize advertising content. In particular, Facebook can track users by using cookies which allow fine-grained targeting for online advertising.

The media and entertainment industry has also been greatly impacted by big data and analytics. Entertainment theme parks (e.g., Disney) use state-of-the-art machine learning techniques to improve their user experience and to increase the profits generated by their parks and by their consumers' derived products (e.g., costumes, mugs, hats, and stuffed toys). In most attraction parks, users can download a smartphone application that allows them to navigate through the park. This application provides the firm access to users' geographic locations in real-time, and thus, allows Disney to better estimate wait times for the different park attractions. This also allows the park restaurants to send targeted offers to specific users, depending on their geographic locations and other attributes.

Traditional restaurants try to use previous data on consumers to improve their service quality and the customer experience. Restaurants often track their customers by requesting phone numbers during reservations. They then record dietary restrictions and preferences in order to customize the service for future visits. At a higher level, restaurants collect data on previous dishes (costs, statistics about people ordering and reordering, online reviews, etc.) and on prices with the goal to constantly improve their offerings and their profits. It even seems that some restaurant owners practice *A/B* tests by varying the menus and prices in order to learn customers' preferences.

Finally, we conclude this section by discussing a recent tool for customer service. Several decades ago, many providers (e.g., banks, telecom companies, and hospitals) opened call centers with the goal of addressing customer complaints and concerns (see, e.g., Aksin et al. 2007, Koçağa et al. 2015). The data obtained from call centers is very large and is often used to learn the problems customers most frequently encounter, and how the company can efficiently address those concerns in a timely manner. Recently, artificial intelligence brought *chatbots* to replace or complement these services (e.g., ReplyYes, Interactbot, and Twyla). A chatbot is a computer program which conducts a conversation via auditory or textual methods to address customers' concerns by appropriately querying the relevant databases (chatbots are often used as part of instant messaging platforms making the interaction with customers convenient). Companies such as Domino's, Pizza Hut, Disney, KLM, Starbucks, Sephora, Staples, Whole Foods, and many others use chatbots to increase customer engagement, promote their products and services, and provide their customers a convenient way to communicate and/or to place orders. Some chatbots use sophisticated natural language processing tools in order to provide a higher quality of service by efficiently scanning databases for matching keywords and word patterns. The presence of large accessible data on previous customers' transactions and complaints has allowed firms in this sector to build enormous datasets. These data sets are routinely used by programs based on artificial intelligence to improve customer service, as well as for information acquisition. For example, a firm can directly ask the chatbot questions about its suppliers, its pending orders, and other matters that can be answered directly from the data.

5. Pitfalls and Risks

The big data revolution raises a number of ethical issues related to privacy, confidentiality, transparency, and identity. Big data brings the natural requirement for issuing laws and regulations on the ways firms can use data, as well as the potential development of big data ethics (Richards and King 2014). This will help protect values such as privacy, confidentiality, and identity theft, as well as avoid illegal discrimination. In this section, we discuss some of the main pitfalls incurred by big data. First, we report examples of recent data leakages that had serious consequences. Second, we highlight several practical challenges related to data accessibility and aggregation. Finally, we consider a serious problem coined *machine bias* that corresponds to the phenomenon in which data-driven algorithms often lead to unfair decisions based on illicit discrimination (e.g., race or gender).

5.1. Data Leakage and Identity Theft

It was reported that 1093 data breaches took place in 2016.[19] One can also count several dozens of large substantial data breaches in the United States in 2017.[20] Two notable examples are Deloitte and Equifax. Deloitte was the victim of a cybersecurity attack in 2017 that went unnoticed for several months. The hackers compromised the firm's global email server through an administrator's account that gave them privileged, unrestricted access. Emails to and from Deloitte's 244,000 staff were stored (some emails had attachments with sensitive security and design details). In addition, the hackers had potential access to usernames, passwords, Internet Protocol (IP) addresses, architectural diagrams for businesses, and health information.

Equifax is one of the three main credit monitoring agencies in the United States that provides credit reports. In September 2017, Equifax announced that the personal data of 143 million US customers had been accessed or stolen in a massive hack. The breach is thought to have revealed the names, Social Security numbers, dates of birth, addresses, and driver's numbers of almost half the US population (44%). Also compromised were the credit card numbers of 209,000 consumers and the personal identifying information of 182,000 users. In addition, the company admitted in October 2017 that the data of some 694,000 British customers was also compromised, some of whom had their financial information and passwords stolen, including partial credit card information. The company's share price plummeted 35% the week after the breach was disclosed (see Figure 2).

This type of data breach can allow hackers to apply for credit cards and loans by stealing the identity of the hacked users. The total losses from identity theft in 2014 amounted to $15.4 billion, with an out-of-pocket loss average of $2895 for the victims.[21] A survey by Gallup News reported that 27% of US adults claim that they were affected by stolen credit card information between October 2015 and October 2016 (up from 22% between October 2014 and October

Figure 2 The Equifax Inc. Stock Price on the NYSE between August 4, 2017, and October 6, 2017 (*source*: investopedia.com) [Color figure can be viewed at wileyonlinelibrary.com]

2015). This increase is partially driven by the presence of larger datasets.

Two additional recent massive hacks were Yahoo and Ashley Madison. Yahoo disclosed in 2017 that all of its 3 billion email users were likely compromised in a 2013 breach, breaking a potential record for the largest ever potential data breach. Yahoo took action and invalidated unencrypted security questions and answers so they could not be used to access an account. Ashley Madison is a Canadian online dating service specializing in extramarital affairs, marketed to people who are married or in relationships. In July 2015, a group of hackers stole user data, by copying personal information, and threatened to release users' names and personal information if the website did not immediately shut down. In August 2015, the hackers leaked more than 25 gigabytes of company data, including user details (containing several thousand corporate emails). This breach received extensive media coverage and had heavy consequences for families, with a few unconfirmed suicides linked to the data breach.

Cybersecurity and privacy are evidently important issues in financial services. Several major operational risk events have happened over the years. A significant breach occurred at JPMorgan Chase in 2014 when information was stolen regarding 83 million accounts.[22] The hackers gained access to the names, addresses, phone numbers, and email addresses of account holders. The bank did not reveal the total cost of the data breach but the bank announced it would spend $250 million a year to improve its cybersecurity. Another event involved the bitcoin exchange Mt. Gox which experienced several security breaches between 2010 and 2014 resulting in losses of approximately $500 million, forcing it to shut down in 2014. Other types of operational risk events that financial institutions have to deal with involve crimes in the form of insider trading, rogue trading (i.e., an employee authorized to make trades on behalf of an employer who makes unauthorized trades), and money laundering. Banks and the US Securities and Exchange Commission (SEC) have put several anomaly detection mechanisms in place to detect such events. These mechanisms are typically based on machine learning techniques, such as Neural Networks, Bayesian Belief Networks, and Support Vector Machines. The presence of big data and online transactions clearly accentuates the risks of rogue trading and money laundering.

Many companies that deal with digital information have several data and cyber analysts who are responsible for detecting fraud (e.g., fraud accounts, advertising fraud, and payment fraud). For instance, it is common to develop data-driven models for cleaning up illicit content from websites (e.g., reviews), primarily based on text and language processing techniques. A second example is models that are based on a set of rules that classify fraudulent and legitimate massive registrations to detect fake accounts. In such models, it is very important to avoid false positives as much as possible. Several startup companies (e.g., Shift Technology) work on developing methods based on artificial intelligence to detect patterns and flag fraudulent insurance claims. This type of algorithm can be trained on hundreds of millions of past insurance claims. Finally, online platforms need to also deal with detecting fraudulent ads (such models are often based on building a dictionary of past fraudulent ads and detecting similarities). It has also become the norm to use multi-factor authentication in order to securely log into many online services. Payment companies such as PayPal invest significant efforts in reducing fraud by developing state-of-the-art algorithms. Such algorithms can be tuned on very large datasets and are typically predictive models that can vary depending on various critical user features. For example, the most valued customers (e.g., monthly users who are high spenders) may face an easier verification process. Such problems are challenging due to the scale of the data and the dynamic aspect.

In the last few years, the field of fraud analytics has exploded. Using data analytics to handle fraud allows organizations to keep control over every transaction. Fraud analytics also identifies hidden patterns and trends to control the exposure of sensitive data across the organization. In addition, it can constantly scan internal systems by trying to detect anomalies in several targeted statistics (e.g., by calculating statistical parameters to find out if or when values exceed averages of standard deviations or by classifying outliers) and learn from previous breaches. A typical method is to use predictive models to compute fraud propensity scores for each transaction in real-time. Then, the firm needs to set the threshold values that dictate the boundary detection for different types of anomalies. Adapting the threshold values dynamically for different users is crucial in order to avoid errors. For more details on data-driven techniques to handle fraud analytics, see, e.g., Bolton and Hand (2002), Delamaire et al. (2009), and Bhattacharyya et al. (2011).

5.2. Data Accessibility, Aggregation, and Acquisitions

Banks and financial institutions have data on millions of customers (and up to hundreds of millions for the largest banks). As a result, the data often needs to be stored in a distributed fashion. Today, most sophisticated banks know how to collect all the relevant data and store it securely. However, one of the main challenges is to make this data easily available to all the relevant users within the institution. Note that these users may range from advanced individuals who can deal with systems and software, such as Hadoop Distributed File System (HDFS), Spark, and Map Reduce, directly to others who have limited coding skills. Many data scientists currently work on such projects with the goal of making the existing datasets accessible to the largest possible number of employees within the institution. This, of course, should be done while being aware of access controls (i.e., sensitive data should be accessible only to a very few users), as well as security concerns. This raises a trade-off between accessibility and security. Several factors can impede having the data easily accessible to a large number of users. One such factor is acquisitions. If an institution acquires another institution (e.g., Bank of America acquired Merrill Lynch in 2008, and Capital One acquired ING Direct USA in 2011), unless a serious effort is invested in integrating the data properly, it will be forever difficult to treat all customers in the same fashion. In addition, there are several regulatory concerns that banks should cope with. One example is the European Union (EU) General Data Protection Regulation (GDPR),[23] which is arguably the most important change in data privacy regulation in the past 20 years. This regulation, which is set to take effect in 2018, dictates, for instance, that algorithmic decisions related to users must be easily explained (the so-called right for explanation) at a later stage.

An additional challenge is to develop efficient ways to use all the data that comes from different channels. Today, customers interact with firms using different modes: offline when customers go to brick-and-mortar locations, online (either via the Internet website or the smartphone application), and through social networks (e.g., many customers report issues with service quality using Twitter). Combining all these interactions and merging the different observations generated by the same user is crucial. Several companies are putting a great effort into this data aggregation endeavor, as ultimately this will give rise to richer and more valuable data. A similar challenge is the one of using private datasets from the firm together with publicly available data (e.g., the NYC Taxi and Limousine Commission database or data from Google Trend). Leveraging the strength of both types of data can allow a better understanding of the market and consumers.

5.3. Machine Bias and Discrimination

In 2017, Propublica, an American nonprofit organization that produces investigative journalism, launched the Documenting Hate project to systematically track hate crimes and bias incidents. It has also used machine learning and natural language processing techniques to monitor and collect news stories about hate crimes and bias incidents. Some of the findings provide evidence for a *machine bias*. As Propublica puts it: "There's software used across the country to predict future criminals. And it's biased against blacks."[24] The organization provided statistical analyses to support this claim and affirmed that basing such decisions on data can be problematic. Recidivism in crimes is only one of many examples. Similar issues arise when insurance companies using data analytics appear to have a bias toward minority neighborhoods that often pay higher car insurance premiums relative to other areas with the same risk. Other typical examples include credit scoring and loan assessment, as decisions in these areas may have ethical and/or legal implications. Consequently, this type of issue raises a flag about being careful when using past data to generalize future insights. The data and algorithms can be biased, and this is not acceptable. This important topic is highlighted in O'Neil (2017) recent book Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. It is shown that in many cases, decision making via mathematical models driven by big data can reinforce discrimination and lead to unfair outcomes. Many researchers in statistics and computer science are working on this problem by trying to ensure the fairness of data-driven algorithms (see, e.g., Calders and Verwer 2010, Hardt et al. 2016). They also address and quantify the bias introduced by data collection, and the issue induced by the fact that many predictor variables can be correlated with the outcome (which we are trying to predict) and some protected attributes (e.g., race or gender). This type of correlation is often referred to as *spurious correlations*, and should be carefully controlled for.

At the end of 2016, an investigation by ProPublica revealed that Facebook not only allows advertisers to target users by specific attributes (such as age, interests, and likes), but Facebook may also let advertisers eliminate users based on race.[25] Facebook released an official statement to defend itself and claims to strictly avoid such practices. As we discussed, the era of big data allows advertisers to target users to an exceptional degree of specificity. However, it is important to train those algorithms to understand which types of attributes are acceptable and which are not. Embedding such knowledge in data-driven algorithms is a clear necessity. The features should be tested for discrimination, fairness, and additional desired requirements depending on the context.

Apart from bias and discrimination, the presence of big data together with digitization allow firms to quickly react. For example, prices on online platforms can vary several times a day and can differ for different users. It is not surprising to observe that two users in the same city are offered different price points for the same product via the same website at the same time. Firms use past data on users' behavior in order to refine their pricing strategies. This gives rise to new issues where users can receive a higher price depending on how often they look at the website, their past searches, the day of the week, the device they are using (mobile versus computer), whether they are using an ad blocker, their geo-localization, etc. Consequently, firms can significantly improve their prediction and profits. At the same time, users can be hurt as they will often get charged a higher price, and this can raise fairness issues. In other words, the fine-grained personalization induced by big data can be perceived as a disadvantage for buyers. A team of researchers at Northeastern University examined 16 popular e-commerce sites (10 general retailers and six hotel and car rental sites) to measure two specific forms of personalization: (i) price discrimination, in which a product's price is user customized, and (ii) price steering, in which the order of the search results is user customized. They found evidence of personalization on four retailer sites and five travel sites, including cases where sites altered prices by hundreds of dollars for the same product. Overall, travel sites showed price inconsistencies in a higher percentage of cases, relative to the control samples (Hannak et al. 2013). It was also claimed that websites such as Expedia and Hotels.com steered a subset of users toward more expensive hotels. It is worth mentioning that in some cases, using big data can actually reduce discrimination. The recent work in Cui et al. (2016) provides evidence for discrimination by hosts against guests of certain races in the marketplace Airbnb. The authors also showed that a review posted on a guest's page significantly reduces discrimination, suggesting that sharing-economy platforms can alleviate discrimination by providing more information and incentivizing peer reviews.

As mentioned in section 3.1, there is a growing recent trend to design experiments that can produce valuable data. Carefully controlled experiments not only attempt to depict the shape of the demand-price curve but also track how this curve changes hour to hour. For example, in some contexts, online purchases may peak during weekday office hours; therefore, retailers are commonly advised to raise prices in the morning and lower them in the early evening. The different deals can vary according to the location, the browsing history, and even the operating system used by the potential buyer. A well-known example is Orbitz which has supposedly targeted Mac users with more expensive search results. Those findings raise the following question: Could the Internet, whose transparency was supposed to empower consumers, be doing the opposite? To alleviate the negative effects of these practices, several tools have emerged to help customers track price changes and detect the best available offers. Examples of such tools are camelcamelcamel.com (a free Amazon price tracker), honey (a free deal-finding browser add-on), and InvisibleHand (a free automatic price-tracker). Those tools offer price history charts and price drop alerts and may also allow users to search for coupon codes whenever they check out online. Such companies generate revenue by earning commissions when users find a sale or a coupon. In summary, the presence of big data allows firms to better price discriminate customers. On one hand, big data can generate higher profits for firms that efficiently exploit historical data. On the other hand, big data can be perceived as unfair by some customers and thus reduce the market share of businesses that use these methods. Finding the right trade-off between these two conflicting effects can be quite challenging.

6. Conclusion

In this study, we discussed how the large amounts of data collected by firms have transformed the service industry. We focused our discussion on services in the following sectors: finance/banking, transportation and hospitality, and online platforms. We presented an overview of how big data has shaped the service industry, discussed several mechanisms that leverage the potential information hidden in big data, and pointed out some of the pitfalls and risks incurred. We conveyed that firms can now collect unprecedented levels of granular data on customers and on transactions. Firms are also developing quantitative data-driven tools to improve operational decisions, such as prices and quality of service. It is clear that having access to large amounts of data can help enhance the service quality by tailoring the offerings to the users' needs.

Combining the power of big data analytics with high-speed computing (which is becoming affordable) allows for real-time service personalization at a very large scale (e.g., online recommendation systems for movies). However, this personalization benefit seems to come at a price. Firms that have access to this rich data can utilize it to price discriminate against customers. In addition, data-driven algorithms can include a machine bias that accentuates illicit discrimination. This raises several legal issues which need to be carefully addressed by governments. Furthermore, the availability of data on sensitive personal information attracts hackers. The number of breaches has increased and is now a major concern for most firms.

Interestingly, there is growing interest in cross-disciplinary services, where many companies try to exploit the interactions between different types of services. For example, Amazon operates in multiple spaces (retail, cloud computing, media streaming, and food delivery services). Airbnb is entering the dining reservation market, and IKEA acquired TaskRabbit, among many other examples.

It seems that having access to big data on different types of services can allow firms to exploit the multi-dimensionality of their users' interactions in order to reach a more comprehensive picture and to enhance the service quality, as well as the long-term profits.

In summary, it is clear that big data has been transforming the way firms interact with customers in the service industry. It is also clear that this transformation is only in its infancy. What is less clear is the extent of the long-term impact of such a disruption. Although big data certainly brings several advantages, some drawbacks are in order. One of the major challenges for firms is to carefully exploit and unlock the power of big data while preserving fairness, trust, and consumers' happiness. Identifying the fine line involved in this trade-off seems to be subtle and may require data scientists, marketers, psychologists, lawyers and regulators to work together.

Acknowledgment

The author thanks Tarek Abdallah, Daniel Guetta, Ilan Lobel, Mike Pinedo, Olivier Rubel, and Luis Voloch for their valuable feedback, which has helped improve this paper.

Notes

[1]https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/big-data-meets-big-data-analytics-105777.pdf.
[2]http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017.
[3]https://www.technologyreview.com/s/514346/the-data-made-me-do-it/.
[4]https://materials.proxyvote.com/Approved/025816/20150313/AR_239749/HTML2/american_express-ar2014_0022.htm.
[5]*Sources*: visa.com, mastercard.com, americanexpress.com, discover.com, 2015.
[6]https://www.federalreserve.gov/newsevents/press/other/2016-payments-study-20161222.pdf.
[7]http://www.experian.com/live-credit-smart/state-of-credit-2016. html.
[8]https://en.wikipedia.org/wiki/List_of_online_payment_service_providers.
[9]http://www.businessinsider.com/venmos-monetization-will-be-worth-watching-2017-1.
[10]https://www.recode.net/2017/7/26/16044528/venmo-8-billion-transaction-volume-growth-rate-chart.
[11]Users can decide to opt for a private mode, where not all the details of the transactions are revealed. However, it was reported that many users keep the default public setting, as they do not bother change the privacy settings.
[12]https://www.economist.com/blogs/schumpeter/2013/01/lending-club.
[13]https://www.lendingclub.com/info/statistics.action.
[14]https://www.etoro.com/.
[15]http://fortune.com/2016/10/20/uber-app-riders/.
[16]https://techcrunch.com/2016/07/18/uber-has-completed-2-billion-rides/.
[17]https://gocurb.com/.
[18]https://www.airbnb.com/about/about-us.
[19]http://www.idtheftcenter.org/2016databreaches.html.
[20]https://www.techworld.com/security/uks-most-infamous-data-breaches-3604586/.
[21]https://www.bjs.gov/content/pub/pdf/vit14.pdf.
[22]https://www.reuters.com/article/us-jpmorgan-cybersecurity/jpmorgan-hack-exposed-data-of-83-million-among-biggest-breaches-in-history-idU.S.KCN0HR23T20141003.
[23]http://www.eugdpr.org/.
[24]https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
[25]This alarming issue was the topic of extensive media coverage, see, e.g., http://fortune.com/2016/10/28/facebook-ad-propublica-race/.

References

Aksin, Z., M. Armony, V. Mehrotra. 2007. The modern call center: A multi-disciplinary perspective on operations management research. *Prod. Oper. Manag.* **16**(6): 665–688.

Amatriain, X. 2013. Big & personal: Data and models behind netflix recommendations. Proceedings of the 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, ACM, New York, pp. 1–6.

Andrews, M., X. Luo, Z. Fang, A. Ghose. 2015. Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Market. Sci.* **35**(2): 218–233.

Arora, N., X. Dreze, A. Ghose, J. D. Hess, R. Iyengar, B. Jing, Y. Joshi, V. Kumar, N. Lurie, S. Neslin, S. Sajeesh, M. Su, N. Syam, J. Thomas, Z. J. Zhang. 2008. Putting one-to-one marketing to work: Personalization, customization, and choice. *Market. Lett.* **19**(3–4): 305.

Bean, R. 2016. Just using big data isn't enough anymore (online). *Harvard Business Review*. Available at https://hbr.org/2016/02/just-using-big-data-isnt-enough-anymore (accessed date July 22, 2017).

Bharath, S. T., S. Dahiya, A. Saunders, A. Srinivasan. 2009. Lending relationships and loan contract terms. *Rev. Financ. Stud.* **24**(4): 1141–1203.

Bhattacharyya, S., S. Jha, K. Tharakunnel, J. C. Westland. 2011. Data mining for credit card fraud: A comparative study. *Decis. Support Syst.* **50**(3): 602–613.

Bimpikis, K., O. Candogan, S. Daniela. 2016. Spatial pricing in ride-sharing networks. Working paper.

Bobadilla, J., F. Ortega, A. Hernando, A. Gutiérrez. 2013. Recommender systems survey. *Knowl.-Based Syst.* **46**: 109–132.

Bolton, R. J., D. J. Hand. 2002. Statistical fraud detection: A review. *Stat. Sci.* **17**: 235–249.

Calders, T., S. Verwer. 2010. Three naive Bayes approaches for discrimination-free classification. *Data Min. Knowl. Disc.* **21**(2): 277–292.

Chen, Y., M. Hu. 2016. Pricing and matching with forward-looking buyers and sellers. Working paper.

Chen, M. K., M. Sheldon. 2016. Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform.

Proceedings of the 2016 ACM Conference on Economics and Computation, ACM, pp. 455.

Choi, H., H. Varian. 2012. Predicting the present with Google trends. *Econ. Record* **88**(s1): 2–9.

Cohen, M. C., R. P. Zhang. 2017. Coopetition and profit sharing for ride-sharing platforms. Working paper.

Cui, R., J. Li, D. J. Zhang. 2016. Discrimination with incomplete information in the sharing economy: Field evidence from airbnb. Working paper.

Cui, R., S. Gallino, A. Moreno, D. J. Zhang. 2017. The operational value of social media information. *Prod. Oper. Manag.* (forthcoming). https://doi.org/10.1111/poms.12707.

Delamaire, L., H. Abdou, J. Pointon. 2009. Credit card fraud and detection techniques: A review. *Bank Bank Syst.* **4**(2): 57–68.

Fong, N. M., Z. Fang, X. Luo. 2015. Geo-conquesting: Competitive locational targeting of mobile promotions. *J. Mark. Res.* **52**(5): 726–735.

Gibson, P. 2006. *Cruise Operations Management*, Elsevier, Oxford, UK.

Golrezaei, N., H. Nazerzadeh, P. Rusmevichientong. 2014. Real-time optimization of personalized assortments. *Management Sci.* **60**(6): 1532–1551.

Hannak, A., P. Sapiezynski, A. Molavi Kakhki, B. Krishnamurthy, D. Lazer, A. Mislove, C. Wilson. 2013. Measuring personalization of web search. Proceedings of the 22nd International Conference on World Wide Web, ACM, New York, pp. 527–538.

Hardt, M., E. Price, N. Srebro. 2016. Equality of opportunity in supervised learning. D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, eds. Advances in Neural Information Processing Systems. *Proceedings of the Neural Information Processing Systems Conference*. pp. 3315–3323.

Hu, M., Y. Zhou. 2017. Price, wage and fixed commission in on-demand matching. Working paper.

Kilby, J., J. Fox, A. F. Lucas. 2005. *Casino Operations Management*. John Wiley & Sons, New York, NY.

Koçağa, Y. L., M. Armony, A. R. Ward. 2015. Staffing call centers with uncertain arrival rates and co-sourcing. *Prod. Oper. Manag.* **24**(7): 1101–1117.

Kohavi, R., S. Thomke. 2017. The surprising power of online experiments. *Harv. Bus. Rev.* **95**(5): 74–81.

Kohavi, R., A. Deng, B. Frasca, T. Walker, Y. Xu, N. Pohlmann. 2013. Online controlled experiments at large scale. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, pp. 1168–1176.

Koushik, D., J. A. Higbie, C. Eister. 2012. Retail price optimization at intercontinental hotels group. *Interfaces* **42**(1): 45–57.

Lucas, A. F., K. P. Brewer. 2001. Managing the slot operations of a hotel casino in the las vegas locals' market. *J. Hospital. Tourism Res.* **25**(3): 289–301.

McAfee, A., E. Brynjolfsson, T. H. Davenport. 2012. Big data: the management revolution. *Harv. Bus. Rev.* **90**(10): 60–68.

O'Neil, C. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books, New York.

Pekgün, P., R. P. Menich, S. Acharya, P. G. Finch, F. Deschamps, K. Mallery, J. V. Sistine, K. Christianson, J. Fuller. 2013. Carlson rezidor hotel group maximizes revenue through improved demand management and price optimization. *Interfaces* **43**(1): 21–36.

Pinedo, M., Y. Xu. 2017. Operations in financial services: Processes, technologies, and risks. *Foundations Trends Technol Inf. Oper. Manag.* **11**(3): 223–242. http://dx.doi.org/10.1561/0200000048.

Richards, N. M., J. H. King. 2014. Big data ethics. Working paper.

Tang, C. S., J. Bai, K. C. So, X. M. Chen, H. Wang. 2017. Coordinating supply and demand on an on-demand platform: Price, wage, and payout ratio. Working paper.

Taylor, T. 2017. On-demand service platforms. Working paper.

Thomas, L., J. Crook, D. Edelman. 2017. *Credit Scoring and Its Applications*. SIAM, Philadelphia.

Xu, Y., M. Pinedo, M. Xue. 2017. Operational risk in financial services: A review and new research opportunities. *Prod. Oper. Manag.* **26**(3): 426–445.

Zhou, R., S. Khemmarat, L. Gao. 2010. The impact of youtube recommendation system on video views. *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, ACM, New York, pp. 404–410.