

INTERNATIONAL JOURNAL OF LAW MANAGEMENT & HUMANITIES

[ISSN 2581-5369]

Volume 6 | Issue 3

2023

© 2023 *International Journal of Law Management & Humanities*

Follow this and additional works at: <https://www.ijlmh.com/>

Under the aegis of VidhiAagaz – Inking Your Brain (<https://www.vidhiaagaz.com/>)

This article is brought to you for “free” and “open access” by the International Journal of Law Management & Humanities at VidhiAagaz. It has been accepted for inclusion in the International Journal of Law Management & Humanities after due review.

In case of **any suggestions or complaints**, kindly contact Gyan@vidhiaagaz.com.

To submit your Manuscript for Publication in the **International Journal of Law Management & Humanities**, kindly email your Manuscript to submission@ijlmh.com.

Big Data: A Boon or A Bane Mystery Revealed

KRITIK KUMAR JAIN¹ AND TUSHAR AHUJA²

ABSTRACT

“Digital- age”, this is the denomination that we proffered to our genesis after closely analyzing our conditioning and positioning it through a variety of tests and scales like we did for stone age or bronze age and exactly like the two periods Internet is the biggest evolution of our times and through it we stay on digital platforms at most of our times which doesn’t charge us a rupee in the conventional sense but we don’t ascertain that nothing in this world is free and what they are charging us is our data which they sell to companies in humongous volumes often entitled as big data, which they put to use to dissect future market tendencies and a lot of other effects like segmenting the merchandise according to market requirements but, there’s a downside to all this too I.e., our data which these companies’ harness is just lying there without any concerns about the information that it could clench. In this study we’re going to talk about some of the operations involved in this process while scrutinizing the benefits and contingencies of big data through various cases, explorations of the industry while suggesting some practices like the Personal Data Protection Bill which is absent even today under the environment of Indian law governance among other effects which we’d suggest in the realm of bigdata and cyberspace.

Keywords: bigdata, forecasting, personal data protection bill, data.

I. INTRODUCTION

Currently in India and in the whole as such we could purview an economic upsurge of people as they ought to persist a lot of money at their disposal, which has prone them to be a bit choosy about *“what they wish to wear, purchase or own”* as a result of which businesses are struggling to keep-up with the current market trends, which had proven to be very dynamic as compared to the past trends.

As back then, we could see that how a particular trend or pattern lasted for at least five years or so, as at first it was adopted by the top-class(richer-class) of the society, then the middle- class took interest over it, and at last the *EBC’s* showed their interest towards it. When it moved from one class to another it generally underwent some minute changes in the pricings in order to

¹ Author is a student at Law College Dehradun, Uttarakhand University, India.

² Author is a student at Law College Dehradun, Uttarakhand University, India.

attract the next class and would work to please the other with a newer design.

But today when a new kind of item rolls out into the marketplace, we can see everyone buying it simultaneously due of which people are getting bored of things which they bought just a few weeks back and this has been the biggest challenge for the production teams of a lot of companies, I.e. how to become the *Status -quo* of the market. One such way is via the mode of *personalizing products* according to the needs, wants and preferences of the consumers and in many fields, this has proven to be the mantra of success. As we could take the example of **Starbucks**, wherein their employees are trained in a manner, so as to provide a positive homely ambience by remembering the names, order preferences etc. of their customers, also Starbucks became a trend-setter by enumerating the trend of “inscription of names” on the coffee(s), which influenced their success ratio to rise humongously and provided them with first movers advantage.

But have you ever wondered how they know so much about you and your tastes and preferences?

Even a new employee you haven't met yet knows your preferred coffee and cake taste, i.e., all of this is feasible simply by putting your information into the company's database, thanks to a procedure known as **Customer Relationship Management (CRM)**.

According to Philip Kotler:

*“Customer Relationship Management (CRM) is the process of carefully managing detailed information about individual customers and all customer “touch points” to maximize customer loyalty”.*³

A pool of data which could be used in aforesaid manner or in forecasting of new trends or in segmentation and varied other ways, and here we are going to study about one of its aspects called Big-Data and how it is both: a blessing and a curse to different people in different spectrums and some of its basic workings.

II. CONCEPT OF BIG-DATA

After discussing its origins and the field of use, now let's discuss about big data as a concept which has revolutionized this whole industry-band, and how they had worked in the past I.e., by a group of experts but now through analyzing heaps of relevant data through statistical data

³ Aashish Pahwa, *Customer Relationship Management: The Ultimate Guide To CRM* (February 12, 2023), Feedough, <https://www.feedough.com/customer-relationship-management-crm-wiki/#:~:text=According%20to%20Kotler%3A,points%E2%80%9D%20to%20maximize%20customer%20loyalty.>

processing processes or through logical reasoning and processes like Artificial Intelligence, a process that has made it more accessible and a lot cheaper.

But before all of these let's see what the basic meaning of big data is, it generally means *“A huge pile of data which can be analyzed in such a way as to gain reasonable insights on current and future market trends or to know more about the behavior of a certain class of customers”*.

“Big data refers to the massive amount of information businesses collect from online and offline sources. These sources include websites, social networks, mobile apps, software, documents, computer logs, sensor networks and more. This explosion of data, however, isn't necessarily significant because of its size but because of what it can do”.⁴

It can help an organization in mainly two forms aside from its other benefits:

- **Forecasting:** One of the most important uses of big data is its use in the process of analyzing and forecasting. As it uses a cornucopia of customer behaviors like, what do they like and what they don't, how their tastes and preferences change overnight, what could be the next big change in their behavioral patterns and what are the factors that might elucidate it and many other things, not only this but we also need to see that the procedure that we opted for while forecasting is moving onto the right direction by taking into account some past data in order to know that it's going in the right way. The best example of this could be “how many fashion brands use this process to know about the latest trends”.
- **Segmentation:** The next phase where this concept has capability to do wonders is Segmentation as under it we can divide or segment our customers into various strata(s) and then provide them what they need, and we could also implement the processes of cross and up selling when a certain class of a segment is getting bored with the brand or products of their choice thereby we provide them with the possible alternatives that could firmly benefit them and hence increase their loyalty towards us which is beneficial for our pockets too or we could even personalize our goods accordingly to cope with the needs and wants of our customers just like, how *Rolls Royce* creates cars in a customized(personalized) manner accordingly as per its customers taste and preferences, they provide about 44000 color combinations to their customers to choose the color(exterior and interior) of their car, that's why it is one of the most entrusted

⁴ Sara Angeles, *Big data and CRM: How Can They Help Small Businesses?*, Business News Daily Staff Writer, <https://www.businessnewsdaily.com/6053-big-data-vs-crm.html>.

luxurious car brand in the world, which starts making its cars only after the order-confirmations.

What really matters is Customer Loyalty, and we can study it through cases of **Coca-Cola**, **Mercedes** and various other companies like for instance, how nowadays in the vending machines, Coca-cola doesn't only imply 120 cans of its soft-drinks, but also allows the consumers to mix-and-match their drinks accordingly as per their requirements and preferences.

- **Competition:** We can not only scrutinize our data, but also can analyze the data of our competitors too to know about what kind of advances they are going to make in the future and how can we counter act those to gain a competitive edge.

III. THE PROCESS

This whole process of big-data has its backbone in the word “**data**” which in today's world we are generating in terabytes, as the famous quote goes “**Data is the new oil**”⁵, though most of it is useless but some of it if looked-upon correctly via *sophisticated data processing tools* could do wonders, generally while researching, we divide such data into two types:

- Primary data
- Secondary data

Here Primary Data talks about a data which is tailor-made to cater to a specific kind of research, and which tends to cater to a specific need, it is generally a bit expensive to generate but it's usage provides us with insightful results which may be highly beneficial in nature, For instance, while conducting a research for footwear products under a specific genre like, sports shoes we shall collect data about the popular games of the region, how deep the pockets of prospects are, what kinds of designs do they prefer and various other aspects. Whereas the Secondary Data talks about huge amounts of pressed data which can or cannot help us after being scrutinized but is available at a cheaper price to work with and comprises of all sorts of matter under its realm, its highly possible that at a casual glance it may all seem to be gibberish after analyzing it through various sophisticated data processing methods it may generate rational acumen.

Next thing before going any further in the process is that we need to analyze the place where this data is stored for the further steps I.e. processing, the terminology used for this is often regarded as “**Data Warehousing**” which is a store-house of a lot of different kinds of data which is to be analyzed and used by the organization. These are generally classified into two

⁵ Humby, C.; Palmer, M., *Data is the New Oil*, (2006, November 3), ANA Marketing Maestros, https://ana.blogs.com/maestros/2006/11/data_is_the_new.html.

kinds: one is operational and while the other is informational here the operational deals with all the day-to-day activities of an organization like, how do a company fares in various situations like, on the meetings and other things taking place while the informational data warehouse handles or stores the data that has to be scrutinized and provides us with the results we need. But the size of warehouses varies from organization to organization as some organizations have big server rooms which contains all these data in external hard drives while the others have divided it into small segments known as **Data Marts** where they are segregated into smaller groups and are often present in smaller quantities when compared to a conventional warehouse. The process of uncovering patterns and other valuable information from large data sets is **Data Mining**, also known as *knowledge discovery in data (KDD)*, and only through this process we can extract the hidden gold(insights) from data through various techniques like **OLAP (Online Analytical Processing)** or through the process of **Normalization** or **Statistical reasoning** or by simply applying if and then rules in the data analysis program. Here the process of Normalization talks about arranging the data logically in a tabular format to minimize duplicity and data anomalies and reinforce the quality and is mainly divided into 5 stages: like 1NF,2NF, 3NF etc. Wherein the first stage has least data integrity while the fifth has least anomaly and is of the sheer quality.⁶

IV. A BOON?

After talking a lot about the whole process up until this stage now let's see if all of it is worth the efforts or not which we could decide through its benefits in various dimensions of works and various case studies to showcase its benefits not merely on the papers but also on the field while talking about the roles big data forecasting dimensions.

(A) Businesses

All the above mentioned examples under this study seems fairly sufficient to determine the benefits of using big data in the businesses, then we would also like to point out the business model of Google and its daughter companies like **YouTube** who are able to keep the lights on and earn a lot of money by just collecting data and using it to generate a truckload of money through their ad revenues a study says google alone generated of about **209.49 billion dollars (about \$640 per person in the US)' worth ad-revenue in 2022.**⁷ This concept of showing ads has not only benefited Google but we would also like to point out that it is also the main area in

⁶ IBM, *What is data mining?*, <https://www.ibm.com/topics/data-mining>.

⁷ Statista(s) team, *Google: Advertising revenue 2022*, Statista, <https://www.statista.com/statistics/266249/advertising-revenue-of-google/>.

which today AI is earning most of its revenue amount.

(B) Manufacturing

The use of machines in the field of manufacturing is also a big point as today it is said that AI (a tool created by processing big data) is going to replace most of the present-day jobs which lack creativity and are generally repetitive in nature. Most of these are in the manufacturing businesses and factories and these jobs are said to be the first to be replaced by machines in the *AI revolution*. As AI works best while doing tasks where we must think inside the box. Moreover, the jobs requiring a creative bent of mind would most probably require a human-touch, people who persist a creative bent of mind, and they will be aided by AI and various big data forecasting tools to make predictions about *what to produce* based on the increase or decrease in order at a particular time in the past.

(C) Scientific Research

The best examples of big-data processing is in the fields of science, wherein we have to proceed a lot of things without even a slightest error in it and with the help of this process we have done a lot of work in fields of biology as it has helped us in *gene-sequencing* to discover a lot of different kinds of proteins while we have done a lot of work in fields of computer and writing software programs, amongst other projects like our *space excursions* is often powered by these processes which are carried out through processing large data sets.

Moreover, these processes help us to advance more and more day in and out.

Whereas under the umbrella of applied and behavioral sciences this process has helped us do a lot of impactful work while creating a lot of new ways to analyze human behavior and discovering new facts and doing a lot of good work to help our society grow even more.

(D) Medicine

As talked about earlier we have used bigdata analysis and AI in gene sequencing and testing of medicines, researchers have claimed that it would even help to cure various lifestyle diseases like cancer and alzymers which we have been struggling to find a cure for more than half a decade and still haven't made a huge dent while finding its cure.

In addition to all this the biggest achievement that it owns is that it has helped in finding the vaccination of the COVID-19 in the fastest way possible as it helped through increasing the speed of various slower processes.

(E) Government

Lastly, the biggest beneficiary of all is the government of a nation, as this process has

revolutionized the way how a government used to function partly through *digitalization*, and which led to storage of data of various citizens in the digital spaces under the realm of various digital databases like the **Digi locker**⁸ and **Ayushman Bharat health accounts**⁹.

This helped the government to access records of its citizens through which they could analyze a lot of things about the key demographic features of a state or could help them to identify the threats to its peace or spying on the anti-national elements.

Although the biggest leap of faith has been taken by the defense sector wherein things like modern drones are doing a lot of the heavy lifting as we could see in the *Russia Ukraine War* wherein a lot of drone strikes are taking place on a daily basis while there are other things too like the modern automated artillery systems among other war machines.

In addition to this the advancements in satellite imagery and photo binning are other things to take note of while they use this process to identify and monitor threats from above without anyone noticing or through facial recognition systems which are used by the US in airports and by their law enforcement agencies where they comb through databases to identify people.

V. VARIOUS CASE-STUDIES

(A) Amazon

The main influence of mine in writing this research article is Amazon, and that how it is utilizing the data of its sellers who sell their products on its website(marketplace) for its usage and through which it is making cheap house copies of these products. As we have seen while searching for a particular item on Amazon we often surf through *amazon's private label offer* which entails the same specifications and a similar look to the corresponding product and is available at half the price, which many people tend to choose because of its low price density and similar looks that attracts the prospects who wish to purchase a cheaper alternative with almost the same qualities as the original one and this roughly amounts to about **3% to 4% of total sales** on Amazon which is in hundreds of billion dollars. Hence, we can see how these private label offers can also account for billions of dollars in sales and because they are generated with the aid of low-quality items and no amount is paid to the producer (seller listed on Amazon) that's why they are often far more profitable to Amazon as compared to the original producers. But all these products that Amazon makes under its private labels are so greatly made because Amazon has their sales records and other data through which they could decide which

⁸ Government of India, *Digi Locker*, <https://www.digilocker.gov.in/>.

⁹ Ayushmaan Bharat Digital Mission, *Ayushman Bharat Digital Mission Creating India's Digital Health Ecosystem*, Ayushmaan Bharat Health Account(ABHA), <https://healthid.ndhm.gov.in/>.

product they need to introduce as their private label and not only Amazon, but various other big supermarket chains duly follow this trend.¹⁰

(B) Walmart¹¹

Another big retail business uses the concept of analysis through big data in its business operations as Walmart uses this process firstly in its *Inventory management system* wherein it utilizes its sales data of previous year to forecast the upcoming years and through this they can even generate predictions about on which items supply they need to double down and at what point of time this helps the company in question a lot as they would never run out of merchandise in their stores and goods in inventory, if they would crack the formula of precisely forecasting what items they need to order for their stores and when.

Not only does this process helps businesses like Walmart, Trends, Big basket and other retail stores in inventory part but it also helps them in *customer retention process* as studies have showed us how one should stick to its old customers as acquiring a new one could cost **10% more** of the company's resources which they could have used in retention of an older one and this could be done by sending text messages to the customers who haven't been regular for a long time and telling them about some new offers and discounts given by the company on the products they used to buy regularly, and which furthermore results in achieving **Customer Loyalty**.

(C) Netflix

Now let's talk about Netflix an online on demand streaming service provider, which occupies more than **one third** of the market in every nation where it operates because it has been successful in cracking and predicting what their customers want to see from its humongous library of movies, TV shows and web-series, through the application of big data as by employing their search history results only, they could predict the behaviors of their users and suggest them what they would love to watch next. According to Netflix, around **75%** of viewer activity is based on personalized recommendations¹². Though it's great for Netflix and its clientele as only through putting into operations the **"search history"** Netflix could show them what would they like to see next but, have we ever questioned to ourselves what could a company know about us if they knew a tad bit more so, here we move on to the next topic of

¹⁰ CNBC, *Secretive Business Of Amazon's 100+ Private-Label Brands*, The Big, <https://youtu.be/8aVslz2oXoA>.

¹¹ DataFlair Team, *5 Big Data Case Studies – How big companies use Big Data*, DataFlair, <https://dataflair.training/blogs/big-data-case-studies/amp/>.

¹² TechVidvan Team, *Top 10 Big Data Case Studies that You Should Know*, TechVidvan, <https://techvidvan.com/tutorials/top-10-big-data-case-studies/?amp=1>.

our research paper I.e., how it is a curse?

VI. A CURSE?

The above mentioned scenario talks about what will happen when a company gets our data and now we are going to talk about things that are happening in the marketplace every now and then, wherein a company is using the data of others to create their own products and one company which has to come to limelight again and again for this very reason is the world's biggest online marketplace which sells its products almost in every country of the world with more than a billion products offered for sale I.e. *Amazon* and as we talked about their private labels and how profitable they are, now we are going to see the other facet of the coin as to gain these insightful data(s) about what is selling at the highest quantity, Amazon uses the data collected by it and creates copies of these brands and sells them under their own brands this is proving to be a very successful venture for them as they can place it above the target which they copied and they also have the power to even place the target at the last page too not to mention how it's sold at a price **20 to 50%** less than the pricing of original product which inevitably decreases their sales and makes the company lose its customer share only because it's selling in the world's biggest marketplace and due to this Amazon has been fined a lot of times by **EU** and has a lot cases filed in **US courts** like in California and others but it doesn't affect them as they are earning in billions and what could a million dollar fine do to them, but while being indulged in malpractices like this Amazon is just affecting its goodwill and flopping its own big sellers and giving other ideas to stay away from it. All the above mentioned things talked about how and what does a company do when it gets our data and next we are going to see the reason why most of the attacks takes place at the HR departments of a company as they are the treasure trove of the data wherein it contains every detail about the people employed to the company and the ones affiliated to them and these attacks today are of a common occurrence through which they take away the data of a lot of their potential clients, current clients, employees, business partners, suppliers and others.

Wherein it contains important details like names, mobile numbers, emails, date of birth, credit or debit card details, pins and passwords, things you like and dislike, travel history among ones sensitive and non-sensitive personal data

Next, we are also going to discuss some examples of recent and old cyber-attacks in order to better understand the data which could get in the hands of the people who intend to sell it on the dark web wherein we would discuss its impact through these case studies.

(A) Aadhar Data Breach Case

One of the biggest data breach not only in India but also the world, the data of **1.1 billion** Indians I.e., personal details were leaked including names, Aadhaar number. Etc., other details that were leaked included identifying information like photos, retina scans and thumbprints among other identifiable details of nearly all Indians and the best thing about it is that our government haven't taken its lesson from this attack, as attacks are happening on the government even today due to their lackluster cybersecurity measures.¹³

(B) Big Basket Data Breach ¹⁴

Here a hacker group named *shiny hunters* delivered the data of around **20 million people** (about the population of New York) was delivered from the company to a dark web forum for **40,000 dollars** (according to cybertheft company who discovered the attack) where it contained email addresses, home addresses, date of births, passwords, Ip-address while interactions taking place in the application. While the group collectively is said to have leaked around **71.3 million** from **11 companies** according to the cycle where the attack was said to have taken place on **14 October 2021**.

In addition to this we have also faced a lot of other attacks like the attack on **HDFC bank** wherein it was alleged that **6 lakh users** were leaked while in the last decade or so there have been a lot of attacks like the one on **yahoo** where it is said the data leaked was of around **3 billion people**(in 2017), whereas on **Facebook attack** leaked data of **533 million users**(in April 2019) moreover the Chinese e-commerce giant wasn't protected either where the data of **1.1 billion users** was leaked (in July 2022) although these attack form the cream of data breaches happening in the last decade many more attacks have been happening on the data centers of companies all over the world some of which are discovered early, while the others remain undetected for a longer period of time .

VII. SUGGESTIONS

To counter these breaches the companies and the government must take stern actions so that the data of their consumers remain protected and untouched while some of them happening could go under the radar, but the magnitude and the sheer number of attacks that are happening today is a big problem for the companies as well as their clients and here are some suggestions

¹³ Abi Tyas Tunggal, *The 70 Biggest Data Breaches of All Time*, UpGuard, <https://www.upguard.com/blog/biggest-data-breaches>.

¹⁴ Anurag, *All you need to know about the Big Basket data breach: Hackers allegedly release data of 20 million Big Basket users for free*, OpIndia, <https://www.opindia.com/2021/04/big-basket-data-breach-hackers-release-data-of-20-million-users>, Last updated (26th April 2021).

which could be used to protect these companies from cases of data breach :

- Human firewall: Under this concept we talk about educating the one who sits behind the systems as in most of the cases of these breaches the root cause is an employee being negligent while doing his work due to which these things happen.
- Putting borders to the borderless cyberspace: The next problem due to which the grip on these attacks can't be tightened is because cyber space is borderless in nature and these attackers could be sitting around anywhere in the world while orchestrating these attacks and most of the times external policies of a nation hinders them to get these defaulters to do justice, thereby the world nations should act in unison to protect the conglomerates and enterprises from the sufferings they bear.
- PDP Bill: This suggestion mainly encompasses to our nation as such, that even in 2023 we lack a Personal Data Protection Bill while most of the huge economies have one to protect the data of their citizens and to find the ones who commit these malice acts and to punish them for their negligent behavior.
- Updating security systems: Next, we'll discuss how we should always upgrade our systems while examining possible vulnerabilities in the code, and that this should be done before the system is employed or infiltrated as a whole.
- International Cooperation: While, in the end, we would suggest that the governments of various nations around the world come together, as they do in the fields of environmental protection, to create some treaties so that criminals from other nations could be easily apprehended and they would act together to put a full stop to these attacks from happening.

VIII. CONCLUSION

Toward the finish of this study as opposed to summing up it in the end, we might want to discuss portion of the patterns that are available in the market today or are getting embraced by organizations consequently we might want to discuss two things in the end: first and foremost, today as opposed to utilizing bigdata some big-tech giants have moved to the idea of *Metadata* which discusses putting away even a bigger measure of information when contrasted with bigdata to utilize, scrutinize and examine while many organizations are as yet utilizing bigdata, while both of them involving similar methodology for its examination.

The next point that we would like to make is that many of these businesses talk about *data privacy*, but some of them don't apply because of clauses in the terms and conditions that no

one reads. Those who resist a clause or two agree to all of them because they can't use the service without agreeing to all the points. And despite the fact that the concept of big data has resulted in numerous new opportunities, such as those for data analysts, we believe that it could pose a

significant problem for the individuals whose data they are processing data that frequently contains a significant amount of our personal and sensitive information and could result in numerous consequences in the event of a data breach.
