Ethics of Big Data Analysis: Theories and Challenges

Introduction

Big data is a digital terminology used to describe huge amounts of data generated by machines or humans. This data is being processed in different fields like hospitality, e-commerce, healthcare management and many others to understand behavior of customers and design future business strategies. However, use of big data also brings a lot of concerns to individual's civil rights in terms of their privacy and security. Big data fosters widespread discussion about its use and sharing it with different stakeholders. This essay will discuss the different concerns encompassing ethical practices in big data analytics and methodologies that can be implemented to raise standards of ethics in big data[1].

In Netherlands on 21 September 2012, a girl from Haren village posted an invitation of her birthday on Facebook. Mass of 3000 people attended the party. As in example it was intended to share with the friends but has reached out to friends- friends as it was publicly shared. Here the ethics come in picture. Ethics for information technology or information system has been studied, but for the big data it is still in the initial stage. Usage of Big data has been increased but the ethical guidelines are still in development. As per Richards and King (3) "large datasets are being mined for important predictions that often yield surprising insights". They claim that all day-to-day life activities (such as shopping, education, cybersecurity, etc.) are impacted by the predictions from Big data; but the individual when in process has very less idea regarding the usage of the data. They note that most revealing personal information such as location history, social network, purchase history, etc. can be mined with the big data. Most of this information is already with government and corporations, and this collection is exponentially increasing. Big data collected is used by the organizations to make variety of decisions for the individual. This problem complicates when the Big data is used to predict and complete the information about the individual based on his/her previous behavior. Hence, compromises the privacy, ownership, and security. To advance in technological aspects with Big data and valuing the security and privacy of the individual; Ethics plays the important role[1,2].

Big data with Ethical theories approach

Ethics is a system of moral principles, or as Richard & Virginia states "Ethics for Big data involves the analysis of conduct that can cause benefit or harm to other people". There are number of ethical theories (such as deontology, utilitarianism, etc.) as it has been studied from long time. To understand the issues related to the Big data from ethical point of view, applying the ethical theories to the Big Data[3].

Kantianism

Kantianism is dutifulness. Kantianism is the philosophy of a German philosopher Immanuel Kant (18^{th} century). Kantian ethics are deontological which means whether it seems right or wrong, it spins around the rule or set of rules (considering humanity and individuals independence). This theory is more concerned about what you should do rather than what you want to do. For instances, one feels right to follow certain etiquettes while few may believe being casual is fine. But under Kantianism every individual should follow same rules. Every person is of the same standard. When the Kantianism is applied to the Big data, it will hard. As, Kantian ethics states that one should always respect independence of any individual and one of the issues with the Big data is privacy. The data by the organizations are collected and analyzed without individual's consent. Organizations may disagree on this, as they specifically mention on website that user data will be used for analysis and will be shared with the partners. And if the user disagrees, he/she can opt-out. Practically, its almost impossible to understand how the data is accessed, shared, and used. It is not only about the individual privacy but also about decisions an organization can make depending on the previous data. Individuals in Big data are considered as data points, which are then used to manipulate in future. Consider an example, a new Netflix user watched a murder mystery movie. For next watch, it will suggest the murder mystery movies. This is not fair with the user, he/she can enjoy some another genre. This is algorithmic manipulation without considering wish of an individual. From Kantian viewpoint, it is questionable whether every individual should approve to a rule that states that the information can be shared with or without their permission, accurate or inaccurate, complete, or incomplete, current, or dated, and can be used with or without their consent. Applying Kantianism is problematic as it believes in maintaining individuality and holding them to same universal guidelines, which in Big data summons the right and fair treatment of individual[3,4].

Utilitarianism

From a social science perspective, the utilitarian approach is to judge the result of a particular act, analysis or a principle based on whether it generates positive or negative effect. For example, a rule or law is right or wrong is whether is generates more net happiness in the society or not, respectively. A right law is one that helps to generate happy feeling in the community if adopted by everybody. Thus, in totality, the right and wrongs need to weigh in to check if a particular rule, law, or an act is morally correct or not, depending on its consequences. Hence, unlike Kantianism where the reasoning behind a particular action or a rule is evaluated, in utilitarianism, the consequence or outcome of particular action critical. Utilitarianism has been followed by media as an ethical principle for a long time now where the cost incurred by a particular action is evaluated against the benefit it incurs to the business of media. There have been many critiques of this principle as the media should not abandon a particular action just because the effects are not in net worth to the initial invested resources. Hence utilitarianism framework for ethical practices in many businesses has not been found to be of morally correct n its absolute terms.

For big data analytics, utilitarian approach is assessed with a utilitarian calculus where good and bad consequences are evaluated and weighed on a scale. The parameters used to evaluate the consequences in big data analytics include intensity of experience of the analytics, its duration,

how probable is a good experience over a bad experience, how close are experience in space and time, to what extent the good experience can get diluted by bad experience and the ability of the analytical approach to create good and bad experience. The analytical approach used to study big data is evaluated for its ethical scale is based on weighing of its benefits and harms to the society. To decide of whether an approach is morally right or wrong is decided by positives and negative effect of the result which helps to select a right alternative with highest net positive effect. The most important limitation of this framework is the ambiguity in deciding the right and wrongs of the approach which can be subjective in many senses and can lead to negative impact on the society. Reaching a consensus on the impacts of a particular big data analytics will be very problematic as costs and benefits will be evaluated on basis of resources spent on analytics and not on social impact. In general, the utilitarian idea is flawed because of the imprecise measurement methodologies and lack of understanding in the society about how the big data affects their lives[3,5].

Social Contract theory

The fundamental philosophy of social contract is to find a way to live in a civilized society and stay free without getting coerced by others in doing acts that we do not agree with. According to the theory proclaimed by Rousseau, social contract is an agreement by which we submit our individual wills to the collective general for a living in society that follows a certain set of rules. Based on the theories proposed by Thomas Hobbs (1588-1679), John Locke (1632-1704), Jean-Jacques Rousseau (1712-1778), and John Rawls (1921-2002), social contract theory is an ethical approach in which any person's moral obligation is based on a contract with other people in a society. According to this theory, people understand that the rationale behind having an agreement within society is to maintain an order to the society without creating any chaos and violating any fundamental rights of an individual. This agreement gives moral and legal backing to the governing system to enforce the rules designated in the contract, for example a government. When applying the social contract theory to big data, the most important role is of regulations created by the government on behalf of the society. It is very challenging to design regulations that will protect the individual privacy rights while still using the data for business purposes in all the different societies. For example, Europe has maintained that the right to privacy is a fundamental right of an individual while the US has different data privacy regulations. In such cases there are negotiations that are required between governments to share and process the data without violating the privacy rights. The key aspect here is different societies envision and articulate the regulations differently and such cases make designing ethical practices for big data a lot more challenging. As big data has become a lot more powerful with the advent of advanced technology in data processing, big data has also become very pervasive, and poses a challenge for the moral societies. Finally, rational people create regulations that seem moral with a view to maintain individual privacy and beneficial for society as a whole[3,6].

Virtue Ethics

Virtue ethics is an approach that emphasizes on the moral character of an individual, and not on rules or consequences as other approaches discussed above. The other approaches discussed above

have a room to give preference to the virtue ethics, but it is not the central theme of the approach. Unlike other approaches, virtue is the foundation of this ethical approach and other normative notions revolve around virtue or moral character of an individual. There have been many philosophers from east that have elucidated on the importance of virtue, but in the west the virtue ethics is designed based on the theories proposed by Aristotle and Plato. Virtue is of two types: intellectual and moral. The intellectual ethics is based on rational thinking and truth. Whereas moral ethics is based on the disposition exhibited by the possessor which is mostly enshrined in the individual due to long term habits that gives a sense of happiness and satisfaction. Virtue ethics relies on the moral character of an individual and his actions. As morality of an individual cannot be left to a set of rules, the virtue ethics examines the character. As big data is not an individual, it is not possible to examine the moral character. So, it becomes important to make sure the intentions of individuals who use and process big data. If an individual is ignoring the fundamental rights of another individual, it is fair to say that the individual is not virtuous to manage big data. While an individual in the health sector is using data to help people in saving lives, the person is considered virtuous. So, depending upon the intentions of the actions of the individual, the ethical framework can be set for big data. Thus, implementing a virtue ethics framework in big data can be very strenuous as one has to be very diligent with examining the intentions of the action of an individual to decide if it is virtuous[3,7].

Ethical challenges in Big data

While applying the ethical theories to the Big data, we can say that there are key challenges to overcome. The data usage from last two decades is increasing rapidly whereas the ethics for Big data is in initial phase and hence, there are more violations. Major industries that fabricate data and are affected by the ethics are Healthcare, education, and information technology. The challenges include Privacy, Security, Ownership, reputation, and evidence-based decision making.

Privacy

Privacy is the freedom from intrusion. In this case, having control over the personal information. Usage of Big data increase as well as the privacy terror continues. Many a times personal information is accessed and shared without the user content. In recent times there were many studies where data was downloaded from social media (twitter) and been used to analyses the COVID situation. In this ideally the shared information was only for the friends or connections, but it was accessed, processed, and evaluated without user's consent. Even though individual in the study may be a data point for the study, yet the privacy is still violated. In the Health industry there are state and federal government codes to protect privacy of the patient. Health Information Portability and Accountability Act of 1996 protects individual health information. It protects data that has identifying information such as names, photographs, etc. Similarly, there are codes for genetic information, adoption, etc. which protects individual. But sometimes when different data is combined in the ways, it threatens the privacy. This data previously may not have invaded any codes, but the combination may scare the individual. As per the cited example, researchers used publicly available information and photographs from Facebook and, with the facial recognition

software, correlated that with the application data with a dating site. In another example, the US health providers distributed the de-identified unknown health information, which can be traced back to individuals with the analytical tool. Hence, with the Big data and the data mining findings (by the correlations among the data) there is always high chances of finding connections based on the similarities. Its not always that the privacy is invade for the negative impact, sometimes it can give the beneficial outcome. For instance, the usage of drug can be tracked. The same can be informed to health officials and make an effort to prevent excessive use of the abusive drug. Similarly, in COVID-19 situation many University Health centers track the patients and try to find the pattern (age, area, etc.) so that public health officials and try control the covid-19 transmission[1,8].

Security

Security is the safety taken to protect an individual. In this case, the individual's data has been protected from the non-permitted sites, organization. More data, there are more data ruptures. Previously, the software or application was limited to the only device or in-house devices. But now, with the cloud, mobile, web-based platforms (social media), it is not restricted to a single location. These platforms are not trustworthy, may prone to the hacking and hence compromise the security of the personal information. The security can be easily compromised on the social media platform. There are many examples of the individuals account on social media been hacked. Sometimes it may have some very serious and adverse effects to your life. Not only on the web-based applications, but there are also few places that demands the information to process further. Such as company having a biometric at the entrance, the company takes the fingerprints to install system but there is no assurance that the prints are deleted after installation. The available security is not always sufficient in protecting the information. To ensure the trust of any individual, security is the important aspect[1,8,9].

Ownership

Ownership is the right of possession. Control of an individual over the data is important thing. Big data is collected from various forum, due to which it is hard to track the data. When data is used the individual should be asked for the consent. Even when an individual sign the consent, they have primary hold over the information. Every organization has the privacy policy, which you must accept. Individual may think it is for the protection, but it is just a liability. Few sites give option of opt-in and opt-out, but that does not change much as individual will still have only the primary level of ownership. The organization have the access to your data and can share it. In Googles case, the email sent from one Gmail account to another are automated to scan and process. But it is applicable only to the individual who have accepted the Gmail privacy policy that means only Gmail users. This automated process applies to all the Gmail as well as other email systems who have not agreed to the privacy policy. Similarly, Facebook was sued for mining the user's private messages for the advertisers. Thus, we can say individual never has ownership of the data[8,9].

Reputation

As per Davis & Patterson, reputation is the significant challenge when it comes to utilization of the Big data. In such problems, its only the organization whose reputation is at stake but its also the users. In case of Target, based on the analysis of clients shopping habit, the company can predict the shopping list and recommends them or gives any offer. In such prediction Target predicted for the clients those are pregnant and the delivery date, so that they can offer the coupons to the mothers-to-be. But in this Target knew that teenage girl was pregnant before her father. Here both the company and client's reputation are at stake. Company claims that it was just the business analytics, but client had a stand saying company is spying on them. As the Facebook and Google's case above shows, even when company are mining the data, their reputation is decidedly negative[1,3,8].

Evidence Based decision making.

To mine the information about the Bis data it uses various algorithms, based on which the decisions are made. As (Tene & Polonetsky 2013) "Predictive analysis may have a stifling effect on individuals and society, perpetuating old prejudices". When predictive analysis is done, and the outcome may not be applicable or appreciated by every individual. In a way it invades the personal space. The decisions are made based on the predictive analysis, even in sensitive topics which can create prejudicial situations. In case of Target, they used predictive analysis on the individuals shopping history to determine the customers who are pregnant. The tactic was to offer the coupons and increase the sell but indirectly the emotions were hurt. What if the customer does not want to reveal her pregnancy? Then it is not ethical just to send the coupons only based on the analysis. Every decision is made solely on the ground of the data. There is still the scope in development of the ethical guidelines[1,10].

Conclusion

Before terminating let us do the evaluation of the Netherland's case. The girl shared an invitation to her friends on Facebook publicly, 3000 strangers showed up. In this scenario the privacy of the girl was invaded. In such situation individuals do not have awareness of power. There should be some changes in social media forms. As of now Facebook provides an individual with many privacy settings, but it has catch. Initially the account is public, which means unless and until he/she are not aware of the right to privacy the account is public and easily accessible. It would be more secure if the Facebook enables all the privacy settings initially and aware individual to explore the different features.

The Big data and the ethical challenges are discussed many times. Ethics are valuable as they tend to help us know whether the argument is right or wrong in the logical way. Big data has become one of the major forces in our day-to-day life. It has all the information about us, what e know about others, what others know about us. We are always the part of it, either we contribute, or the devices connected contribute. Here, we tried to apply the ethical theories to the Big data analysis to understand the differing perspective of the moral situations. Concluding about these theories

and challenges we can say that Virtue Ethics is the theory that can be one of the fits. Application of theories are intended to evaluate the course of action for any individual and organization without any challenges. Few challenges we have discussed which needs to be focused to avoid any trouble to individual or organization.

References

1. Zwitter, A. Big Data ethics. *Big Data Soc.* **1**, 1–6 (2014).
2. Richards, N. & King, J. Wake Firest Law Review. *Big Data Soc.* **49**, 2053951714559253 (2014).
3. Herschel, R. & Miori, V. M. Ethics & Big Data. *Technol. Soc.* **49**, 31–36 (2017).
4. Kant, I. Kantian Ethics. *https://www.csus.edu/indiv/g/gaskilld/ethics/kantian%20ethics.htm*.
5. Utilitarian Theories. *http://caae.phil.cmu.edu/Cavalier/80130/part2/sect9.html*.
6. Celeste Friend. Social Contract Theory. *Internet Encyclopedia of Philosophy*.
7. Virtue Ethics. *Stanford Encyclopedia of Philosophy* (2003).
8. Wielki, J. Social Ethical Challenge Tech. *Polish J. Manag. Stud.* **11**, 192–202 (2015).
9. White, G. Issues in information systems big data and ethics: Examining the grey areas of big data analytics. *Issues Inf. Syst.* **17**, 1–7 (2016).
10. Saqr, M. Big data and the emerging ethical challenges. *Int. J. Health Sci. (Qassim).* **11**, 1–2 (2017).