Data Ethics: The Moral behind Data

Dr. Samiksha Shukla
Associate Professor & HOD, Department of Data Science
CHRIST (Deemed to be University), Lavasa, Pune
samiksha.shukla@christuniversity.in

Mr. Kapil Tiwari
Associate Director, Product Development, SS&C Eze Software, Hyderabad
ktiwari@ezesoft.com

Introduction

Data Science is one of the most prominent emerging domain in the world these days. To make life comfortable and smooth various researches are going in and around Data Science. One very important aspect comes with this is ethics. Data should be handled ethically, if not, leads to furthermore issue for individual and organization. At present with the rapid societal and technological revolution, notions like 'confidentiality,' 'impartiality,' and 'representation' are redefined. We are living in an era where fake news propagates at a faster rate than the truth, our societies are at an acute moment. At this juncture, we need to have discussions about our individual and mutual responsibility to handle data ethically.

Data Ethics (DE) is treated as one of the most important challenge of moral problems related to data. As, Data is generated and utilised at significant rate of 2.5 quintillion records every day, handling it carefully is very important. Individuals are being defined by how & where they eat, commute, use Internet and time span of all the online activities. Society is in the midst of a "data revolution," where individuals and organizations can stack and analyse huge amounts of data. Data Analytics can bring across amazing findings and innovations which can empower the society: from applying machine learning to healthcare research to binding data to build smart cities.

An essential aspect of DE is, to use Data Science correctly. Enormous prospects are provided by Data science for betterment of private and public life. Unfortunately, such prospects are also connected to substantial ethical challenges. The size of data (mostly private) is continuously growing. The users are relying on recommendation algorithm to investigate the data. This investigation helps users to choose appropriate products from a variety of available products. The recommendation system is the major ethical challenge faced by an individual, as one may not be willing to share his or her online behaviour. This leads to gradual reduction of human involvement but raise concern related to accountability and adherence to the human rights, as sharing private data is unethical.

Appropriate use of data presents various opportunities to provide finest possible services to the users. Though these new opportunities come with new challenges, DE plays key role in resolving these challenges. Individuals and organizations are bound to abide with the principles of DE. There are various legal obligations imposed, when it comes to DE. These obligations differ with the type of data and the source from where it is received. Higher Education Institutes are including DE in curriculum to coach the younger generation of the country to understand importance of DE and obey it.

Data Ethics Outline

The purpose of this outline is to create responsible data usage culture in the society where social media practise is at its peak. Figure 1 shows the authors view on the facts which need to be balanced by users concern with DE. In the figure a scale is shown, balancing the facts to maintain the DE. Information technology providers can understand the core ethical obligation of the data involved and propose their offerings appropriately.

- **Ethical Practices:** Now a days it is mandatory for organizations to train the employees on General Data Protection Regulation (GDPR), even governments are taking initiative to train the citizens about ethical data usage.

- **Moral Values:** No matter which country you belong to, the moral value associated with data will be similar based on the domain and complexity of the data. Primary moral value of the data is privacy, followed by ownership, trust and respect.

- **Code of Conduct:** Code of Conduct is decided by the government and the organizations, based on what is considered as appropriate and inappropriate for them. It may be possible, what is appropriate for one may not be the same for the other.
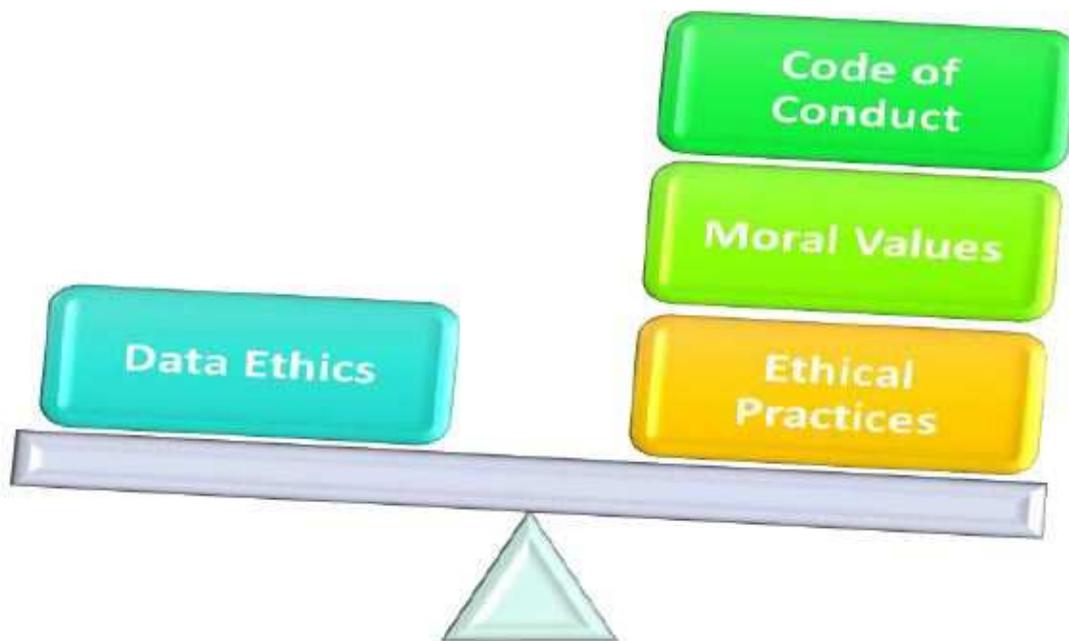
Figuer 1: Component of Data Ethics

Need of the Data Ethics

Responsible usage of data is the need of the hour. According to the study conducted in 2019, there are 3.48 billion social media users with an increase of an average of 8% every year. The social media user population has radical increase but most of the users are unaware of the DE principles. World's internet users spend on an average 6.5 hrs a day online. Google is the most visited website with first rank, followed by YouTube and Facebook.

Today's technology is utilised at a faster rate due to the influence of social media and automation everywhere. In this era of Data, it is very important to have set moral rules for DE. DE is required pertaining to creation, storing, managing, processing, distribution and utilisation of data. DE is also required for the algorithms and resultant practices such as programming, hacking, professional codes and innovation. DE is the need of the hour in order to devise and endorse ethically respectable solutions.

As per the study conducted by DataReportal [6], in Indian context for the total population of **1.361 billion** the mobile subscriptions are **87%, 41%** population uses Internet and **23%** population are active social media users. The active social media users have increased by **25%** in India as compared to the year **2018**. In the Indian context average online time per user is **7 hr 47 m, 89 %** of Indian internet users prefers mobile messengers, with **WHATSAPP** and **FACEBOOK** as most preferred apps. Considering the kind on usage catered by the Indian population there is a need for a set of rules to be followed so that the individuals can contribute their share of DE.

Data Ethics Characteristics
These DE Characteristics is presented to encourage ethical data usage on the core values of the Public Service Code of ethics like integrity, honesty and objectivity - in order to deliver better services and information policy. It will present authors view about the rules and the principles regulating the usage of the data by an individual or an organization.

The speed at which technology is changing, we need to monitor that the moral issue related to the data are addressed considering ethical value of data usage. It requires suitable guideline which will help data practitioners to use data responsibly without any ethical issues. The proposed guideline will help the government, policy makers and the public sector to handle data responsibly and understand the process to support the ethical usage of data. As diverse data is utilised by individual and organization for different purpose, many ethical challenges are encountered. These challenges always comes with data, but still there is no black and white rules which everyone will agree upon.
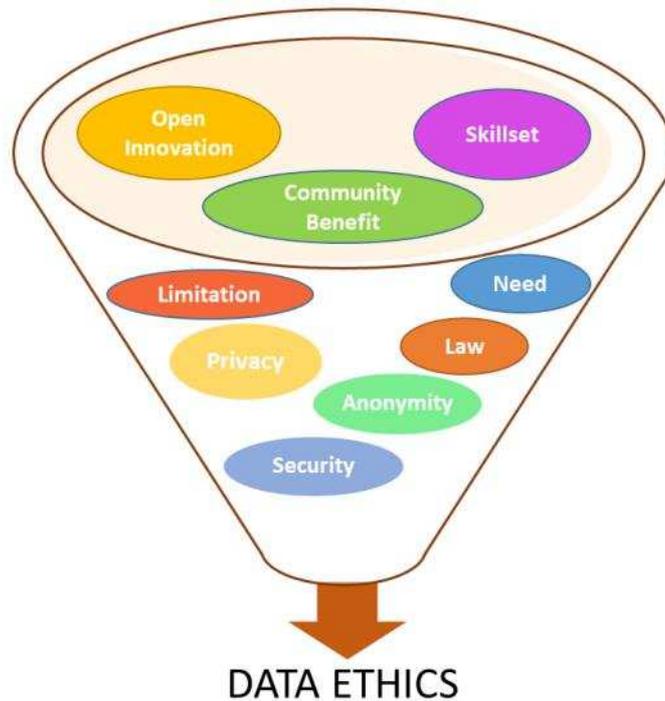
Figure 2: Data Ethics Characteristics

The key features shown in **figure 2** includes various elements which needs to be considered when working with the data like Skillset, Open Innovation, Community Benefit, Privacy, Anonymization etc.

Data Ethics for Individual

The data can be utilised by individual for personal research & innovation, individual growth, business forecasting & prediction. As a responsible citizen we should follow certain DE such as:

Key 1: Support vigorous practises and consider your expertise
Individual should work within their skillset recognising where the domain expertise is required so that, knowledge transfer can be planned. Awareness is the key for individual such as knowing about data, consistent practices e.g. multidisciplinary team, expert opinion, accountability of algorithm, testing under different conditions, reproducible design.

Key 2: Support Open Innovation and Open up Opportunity for all
Individuals involved in design and development should be transparent about the tools, data and algorithms used to achieve the goal. This helps other researchers to inspect the findings and appreciate the novelty of the work.

Data Ethics for Public and Private Sector

Whenever there is a requirement to follow DE principles by public service providers or data practitioners the following key points should be considered.

Key 1 - Begin with the data prerequisite and the community benefit
The data when utilised in effective way has the potential to transform how public services are delivered. We must always be clear about what we are trying to achieve for users by the data available to provide better services. Need and expected public benefit analysis is prerequisite before starting the work with the data. It will help data users to take right approach. Clear user needs help the government and organization to understand the problem. Considering the DE Principles holistically help us ensure public benefit.

Key 2 - Awareness about Legal challenges and Code of conduct:
Government across the geographies are in the process of making strategies for data usage. Organizations and service providers should abide the law followed by the judiciary. To obey this an organization must understand the significant laws and codes of conduct which is applicable to the type of data usage. In case of doubt expert opinion is sagacious.

Key 3 - Use Data as per Need
The data must be used in proportion to the user requirements. Only necessary data should be used which is required for attaining desired outcome. Over usage of data may slowdown the processing and increase the waiting time to achieve the

desired result. In order to understand that the intended data is proportionate to the need, officials should consult public or take advice from ethical committees and experts.

Key 4 - Understand the limitations of the data

Different type of data have different set of limitation imparted for the data sharing and usage. Government should clearly mention the data sharing and usage policy for different sector like civil services, health service, population, social etc. It is the responsibility of a government official to use data considering its limitation.

Conclusion

With the use of data comes the misuse of the data. DE is essential when we talk about individual, private or public sector. DE principles helps us understand different perspective of right and wrong usage of the data. In any case, where data is involved prior to initiating the work with the data, we should always consider the user need and community benefit. Clear user need gives deep understanding of the problem. Many public services can use data analysis to operate and improve the services, provide new services, testing new and existing policies. So that, it can help officials to identify the impact of new policies, area of improvement and the needs as per the geographic patterns.

All the officials and individuals involved in data practices must be aware of legal challenges and code of conduct implicated to data usage. The Personal Data Protection Bill, 2018 (PDPB) [1] ensures protection of individuals' personal data and regulates the collection, usage, transfer and disclosure of the data. PDPB clearly states the legislation related to the different type of data, like personal, personal sensitive, data portability, cross border transfer of personal data etc. The Data practitioners and Leaders often speak about the questionable and illicit sharing, collection and usage of sensitive data. To drive change and hitch the positive impacts of the data, we need collective effort. These efforts need to reach beyond academia and industry to general public, to make this world a better place to live. These conversations should not be in symposia, conferences or workshops but also over the dining tables everywhere.

The discussion in the article is authors view about data ethics.

References:

1. https://www.meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill,2018.pdf
2. https://www.datascience.com/blog/data-ethics-for-data-scientists
3. https://datafloq.com/read/big-data-ethics-4-principles-follow-organisations/221
4. http://theconversation.com/data-ethics-is-more-than-just-what-we-do-with-data-its-also-about-whos-doing-it-98010
5. Zwitter A (2014) Big Data ethics. Big Data & Society1(2):2053951714559253
6. https://datareportal.com/

About the authors

Dr Samiksha Shukla, is currently employed as Associate Professor and Head, Data Science Department, CHRIST (Deemed to be University), Lavasa, Pune Campus. Her research interest includes Computation Security, Machine Learning, Data Science and Big Data. She has presented and published several research papers in reputed journals and conferences. She has 15 years of academic and research experience and is serving as reviewer for Inderscience Journal, Springer Nature's International Journal of Systems Assurance Engineering and Management (IJSA), and for IEEE and ACM conferences. https://www.linkedin.com/in/samiksha-shukla-phd-14ba93a/

Kapil Tiwari is currently employed as Associate Director, Product Development SS&C Eze Software, Hyderabad. He has overall 15+ Years of professional experience in Software Design and Development using C++, C#, AngularJS, NodeJS, COM, with organization like OpenText, EMC, Sun Microsystems and Siemens information system ltd. Kapil also has rich exposure of Finance Technologies, Content Management and Medical domain. He is serving as program committee member and reviewer for IEEE and Springer Conferences. He has designed and developed products into Microsoft Office, Office 365 and SharePoint add-in space. https://www.linkedin.com/in/kapiltiwari1981/

"You can have data without information, but you cannot have information without data." – By Daniel Keys Moran

"Torture the data, and it will confess to anything." – By Ronald Coase

"Data really powers everything that we do." – By Jeff Weiner, CEO of LinkedIn