



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)



Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai
Accredited by NAAC-UGC with 'A++' Grade (Cycle III) & Accredited by NBA (B.E - CSE, EEE, ECE, Mech & B.Tech.IT)
COIMBATORE-641 035, TAMIL NADU

DEPARTMENT OF COMPUTER APPLICATIONS

19CAE716 – DATA SCIENCE

UNIT – I: INTRODUCTION TO DATA SCIENCE

**TOPIC: DATA SCIENCE ETHICS & DOING GOOD DATA
SCIENCE**



Introduction

- ✓ Data may be used to enhance judgments and significantly impact the company.
- ✓ Business people who handle data must adhere to specific ethical standards.
- ✓ Data contains personal information about individuals, and thus it must be used appropriately to protect privacy.
- ✓ The ethics in data science must be a concern for analysts, data scientists, and IT professionals.
- ✓ Anyone who deals with data must be familiar with the basics. Any incidence of data theft, storage, unethical data collecting, usage, etc., must be reported by anybody working with any form of data.





Data Science Ethics

- ✓ The study and evaluation of ethical problems connected to data have given rise to a new field of study in ethics, known as ethics for data science.
- ✓ Data may be collected, recorded, produced, processed, shared, and used, among other things. It also encompasses different data and technology, such as programming hackers, professional codes and algorithms.
- ✓ Data ethics builds on and expands the boundaries of computer and information ethics. They are shifting from being information-focused to being data-focused.
- ✓ Many ethical questions are raised about the data businesses acquire from ordinary people. This is becoming more crucial as companies start to monetize the data, they have gathered from individuals for uses other than those for which it was initially obtained.





Importance of Data Ethics

- Data science significantly influences how industries do business. The risks of data science without ethical considerations are as evident as ever. When appropriately used, algorithms have a great deal of potential to improve the world.
- The advantages might be significant when we use robots to carry out tasks that previously needed a person.
- The importance of ethics in data science has been felt because there has to be a clear set of rules governing what businesses can and cannot do with the personal information they acquire from customers.
- All experts usually agree that some basic things should be implemented, even if there is still a lot of grey space in this field and nothing is black and white.
- These are only a handful of the significant subjects and guiding ideas now receiving the most attention. However, there is still a lot of space to cover regarding the creation and development of data ethics.



Principles of Data Ethics

- ✓ The word privacy does not imply confidentiality since private information may be required for audits depending on the needs of the legal process. However, this sensitive information is obtained from a person with their permission. Additionally, it is stated that the information must not be made public so that other people or businesses might use it to determine the user's identity.
- ✓ Private information that has been disclosed should never be made public. In order to protect the privacy and comply with regulations, they must also impose limitations on how the data may be shared.



**BIG DATA
BIG MONEY**



- ✓ Customers should be open-minded about how the data is being used or sold. Additionally, they must be able to manage the transfer of their data across independent, enormous analytical platforms.
- ✓ Big data should not interfere with human will in any way. Because big data analytics may ascertain and even influence who we are before making decisions. It is one of the ethical criteria for utilizing the ethics of data analytics.
- ✓ Big data shouldn't institutionalize prejudiced attitudes like sexism and racism are common examples. Machine learning algorithms can strengthen people's unconscious prejudices through many training examples.



Doing good Data Science



Privacy:

Data Collection: Obtain informed consent for data collection, especially when dealing with personal or sensitive information.

Data Minimization: Collect only the data necessary for the intended purpose.

Anonymization: De-identify or anonymize data where possible to protect individual privacy.

Transparency:

Communication: Clearly communicate data practices, including how data is collected, used, and shared.

Algorithms: Be transparent about the algorithms and models used in data analysis.

Fairness and Bias:

Fair Data Practices: Ensure that data collection and analysis do not result in discriminatory practices.

Algorithmic Fairness: Evaluate and address biases in algorithms to avoid perpetuating or amplifying existing inequalities.

Informed Consent:

User Understanding: Ensure that individuals understand how their data will be used and for what purposes.

Revocable Consent: Provide mechanisms for individuals to revoke consent and manage their data preferences.



Doing good Data Science



Data Security:

Secure Storage: Implement robust security measures to protect data from unauthorized access.

Data Encryption: Encrypt sensitive data to prevent unauthorized interception

Data Ownership and Access:

Respect Ownership: Acknowledge and respect the ownership of data, whether it belongs to individuals or organizations.

Controlled Access: Restrict access to sensitive data to authorized personnel only.

Data Quality:

Accuracy: Ensure the accuracy and reliability of data used in analysis.

Account for Biases: Be aware of biases in the data and take steps to mitigate their impact.

Accountability:

Ethical Oversight: Establish ethical review processes for data science projects.

Responsibility: Hold data scientists and organizations accountable for the ethical implications of their work.



Doing good Data Science



Continuous Monitoring:

Ongoing Assessment: Regularly assess the ethical implications of data practices throughout the data lifecycle.

Update Policies: Adapt ethical policies and practices to changing technologies and societal expectations.

Societal Impact:

Consider Broader Implications: Evaluate the societal impact of data science projects, including potential unintended consequences.

Social Responsibility: Take into account the ethical implications of decisions on a broader scale.

Professional Integrity:

Honesty: Be transparent about limitations and uncertainties in data analysis.

Avoid Misrepresentation: Do not manipulate data or results to fit a particular narrative.

Data Governance:

Establish Policies: Develop and enforce data governance policies that align with ethical principles.

Training: Provide training to data science teams on ethical considerations and best practices.



Doing good Data Science



Legal Compliance:

Adherence to Laws: Ensure compliance with data protection and privacy laws applicable to the region or industry.

Stay Informed: Stay updated on changes in regulations and adapt practices accordingly.

Collaboration:

Interdisciplinary Collaboration: Work collaboratively with experts from diverse fields, including ethicists, legal experts, and domain specialists.

Public Engagement:

Engage with Stakeholders: Involve stakeholders, including the public, in discussions about data practices and ethical considerations.

Feedback Mechanisms: Establish mechanisms for receiving and addressing public concerns about data use.

