Responsible analytics practices

The meaning of "Responsible analytics practices" are a set of principles and guidelines that organizations should follow to ensure that they are using data in a way that is ethical, fair, and transparent. These practices are important to protect the privacy and rights of individuals, and to build trust with customers and stakeholders.

5.1 Data privacy laws and best practices: The meaning of" Data privacy laws and best practices" are Data privacy laws, also known as data protection laws or regulations, are legal frameworks that govern the collection, use, processing, storage, and sharing of personal data. These laws are designed to protect individuals' privacy by ensuring that organizations and entities handle their personal information responsibly and securely. Data privacy best practices are a set of guidelines and strategies that organizations and individuals can follow to protect personal information and comply with data privacy laws. These practices aim to ensure that data is handled with care, respecting the privacy rights and expectations of individuals. Here some keys point of Data privacy laws and best practices:

1.Data Protection Regulation (GDPR)

- Definition: General Data Protection Regulation (GDPR) is a regulation in the European Union that sets out strict rules for the collection and use of personal data. The GDPR applies to all organizations that process personal data of individuals located in the European Union, regardless of where the organization is located.
- Example: A company that collects data from customers in the European Union must comply with the GDPR. This means that the company must obtain consent from customers before collecting their data, and it must provide customers with access to their data and the ability to correct any inaccuracies.
- Image example: https://www.loginradius.com/compliance-list/gdpr-compliant/

2.Family Educational Rights and Privacy Act (FERPA)

- Definition: Family Educational Rights and Privacy Act (FERPA) is a US law that protects the privacy of student education records. FERPA gives parents and students certain rights over their educational records, including the right to access, correct, and delete their records.
- Example: A school district that collects data from students must comply with FERPA. This means that the school district must obtain consent from parents before collecting data from their students, and it must provide parents with access to their students' data and the ability to correct any inaccuracies.
- Image example: http://news.unm.edu/file/FERPA%20logo?action=

3.Health Insurance Portability and Accountability Act (HIPAA)

- Definition: Health Insurance Portability and Accountability Act (HIPAA) is a US law that protects the privacy of health information. HIPAA applies to healthcare providers,

healthcare insurance companies, and other organizations that handle health information.

- Example: A doctor's office that collects data from patients must comply with HIPAA. This means that the doctor's office must obtain consent from patients before collecting their data, and it must protect the data from unauthorized access or use.
- Image example: https://www.whoa.com/how-you-can-avoid-hipaa-violations-in-the-cloud/hipaa-logo/

4.Institutional Review Board (IRB)

- Definition: Institutional Review Board (IRB) is a committee that reviews research proposals to ensure that they protect the rights and welfare of human subjects. IRBs are required to review research proposals that involve human subjects, including research that uses personal data.
- Example: A university that is conducting research on student behavior must obtain approval from an IRB before collecting data from students. This ensures that the research is conducted in a way that protects the privacy and rights of the students.
- Image example: https://34.195.119.223/logo-irb-brasil/

5.Payment Card Industry Data Security Standard (PCI DSS)

- Definition: Payment Card Industry Data Security Standard (PCI DSS)is a set of security standards for organizations that process credit and debit card payments. The PCI DSS requires organizations to implement certain security measures to protect credit and debit card data from unauthorized access, use, or disclosure.]
- Example: A merchant that accepts credit and debit cards must comply with the PCI DSS. This means that the merchant must implement security measures to protect the credit and debit card data that it collects from customers.
- Image example: https://getlogovector.com/pci-dss-compliant-logo-vector-svg/

5.2 Describe Best Practices for Responsible Data Handling:

The meaning of "Describe best practices for responsible data handling" is to explain the best practices in responsible data handling. These best practices encompass the most effective and ethical ways of collecting, storing, securing, interpreting, and using data. The purpose of these practices is to protect individual privacy, ensure data security, prevent data misuse, and ensure that data is used in a legal and beneficial manner. This involves compliance with applicable data privacy laws and ethical guidelines in data usage. Here some keys point of Describe Best Practices for Responsible Data Handling:

1. Methods of Handling PII (Personally Identifiable Information):

- Definition: PII refers to information that can be used to identify individuals, such as names, addresses, ID numbers, or phone numbers. Best practices involve strict protection of PII, including data encryption when stored and transmitted.
- Example: E-commerce Shopee should encrypt customer information, such as social security numbers, when it's stored in their databases. Access to this data should be restricted to authorized personnel only.

- Image example:
  https://slack.com/trust/security
  https://seeklogo.com/vector-logo/326282/shopee

2. Securing Data:
   - Definition: Data security involves using encryption, firewalls, limited access, and active monitoring to protect data from unauthorized access.
   - Example: MyEduSolve should use encryption to protect sensitive financial data during transmission, ensuring that even if intercepted, the data remains confidential and secure.
   - Image example:
     https://www.iconfinder.com/icons/4047377/data_fire_firewall_firewalls_network_security_server_icon
     https://images.app.goo.gl/sE6yHVjagModuuuu5

3. Protecting Anonymity Within Small Data Sets:
   - Definition: When working with small datasets that may have the potential to identify individuals, it's essential to carefully design the data to maintain anonymity.
   - Example: Researchers working with a small survey dataset might aggregate responses for demographic information, making it harder to identify any single participant.
   - Image example: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1125011

4. Importance of Anonymizing Data:
   - Definition: Anonymization is the process of removing information that could identify individuals from data. This is important for safeguarding privacy and complying with regulations such as GDPR.
   - Example: A Data Analytics before sharing healthcare records for research purposes, all personally identifiable information, like names and addresses, should be removed to protect patients' privacy.
   - Image example:
     https://www.webstaurantstore.com/complyright-hipaa-protecting-patient-privacy-poster/529A2126.html

5. Trade-Offs When Balancing Interpretability and Accuracy:
   - Definition: There are situations where you must choose between maintaining data interpretability and achieving higher accuracy. For example, very complex machine learning models may offer high accuracy but are harder to interpret.
   - Example: A Data Analytics or Data Scientist In credit scoring, a simple model that uses just a few variables may be more interpretable, even if it's less accurate, while a complex neural network might have higher accuracy but is less transparent.
   - Image example:

https://towardsdatascience.com/the-balance-accuracy-vs-interpretability-1b3861408062 or
https://www.nature.com/articles/s41598-022-05079-0/figures/1

6. Shortcomings of Making Population-Level Generalizations with Limited Sample Data:
   - Definition: When making generalizations about an entire population based on a limited sample, there are risks of error. The sample may not fully represent the variation within the population.
   - Example: the share of U.S. adults who said they had received at least one COVID-19 vaccine dose by June 2021 was roughly two-thirds based on data from both the Centers for Disease Control and Prevention (66%) and Center polling (67%). While not perfect, this level of accuracy is usually sufficient for getting a meaningful read of the public's mood on key issues.
   - Image example: https://www.pewresearch.org/?attachment_id=415256

5.3 Given a scenario, describe types of bias that affect collection and interpretation of data: It is important to recognize and address these biases in data collection and interpretation to ensure the integrity and validity of the findings. This can involve careful study design, transparent methodology, and rigorous data analysis techniques to mitigate bias as much as possible.

1.Confirmation bias
(is the tendency to search for, interpret, favor, and recall information in a way that confirms or supports one's existing beliefs or hypotheses)
   - Definition:
     Confirmation bias is a type of cognitive bias that occurs when we only seek out. information that confirms our existing beliefs, while ignoring or downplaying information that contradicts them.
   - Example: A person who believes that climate change is not real may only seek out information from sources that support their belief, while ignoring or downplaying information from sources that contradict it.
   - Image example: https://www.bbc.co.uk/bitesize/topics/zw982hv/articles/z4mts82

2. Human cognitive bias
(is a mental shortcut that can lead to errors in judgment. Cognitive biases are based on our experiences, emotions, and assumptions.)
   - Definition: Human cognitive bias is a mental shortcut that can lead to errors in judgment. Cognitive biases are based on our experiences, emotions, and assumptions.
   - Examples:
     Availability bias: The tendency to give more weight to information that is more readily available.
     Anchoring bias: The tendency to rely too heavily on the first piece of information we receive when making a decision.
     Representativeness bias: The tendency to judge the likelihood of an event based on how similar it is to other events we are familiar with.

- Image example: https://www.smithsonianmag.com/history/slavery-trail-of-tears-180956968/

3. Motivational bias
(is a type of bias that occurs when our motivations influence our collection or interpretation of data.)

- Definition: Motivational bias is a type of bias that occurs when our motivations influence our collection or interpretation of data.
- Examples: A researcher who is trying to prove a particular hypothesis may collect data in a way that is biased towards supporting their hypothesis. A company that is trying to sell a product may present data in a way that is biased towards making the product look more appealing.
- Image example: https://www.frontiersin.org/articles/10.3389/fdata.2022.787421

4. Sampling bias
(occurs when the sample of data collected is not representative of the population that it is intended to represent.)

- Definition: Sampling bias occurs when the sample of data collected is not representative of the population that it is intended to represent.
- Examples: A poll that only surveys people who are registered to vote is likely to be biased towards the views of registered voters. A study that only looks at people who are hospitalized with a particular disease is likely to be biased towards the views of people who are hospitalized with that disease.
- Image example: https://www.shopify.com/id/blog/206934729-how-to-start-a-clothing-line

5. Selecting visualizations/data representations to avoid bias

- Definition: When selecting visualizations and data representations, it is important to be aware of the potential for bias. Some visualizations and data representations can be more misleading than others.
- Here are some tips for selecting visualizations and data representations to avoid bias: Use visualizations that are accurate and easy to understand. Avoid using visualizations that are misleading or manipulative. Be transparent about how the data was collected and analyzed. Use multiple visualizations to represent the same data, and compare the results.
- Example: Definition: Using a bar chart to compare two groups of people is a good way to avoid bias, because bar charts are easy to understand and do not distort the data. However, using a pie chart to compare the same two groups could be misleading, because pie charts can make small differences appear larger.
- Image example: https://www.quora.com/What-is-the-difference-among-histogram-bar-chart-pie-chart