Data Analytics Project

24. Bias & Ethics in Data Analysis

Aagam Deolasi

1

WHAT is Bias?

Data bias occurs when an information set is inaccurate and fails to represent the entire population. It is a significant concern as it can lead to biased responses and skewed outcomes, resulting in inequality. Hence it is important to identify and avoid them promptly.

Aagam Deolasi

2

TYPES of Bias

Following are the common types of Biases in Data Analysis:

A. Confirmation Bias
B. Selection Bias
C. Interpretation Bias
D. Information Bias
E. Predictive Bias

Aagam Deolasi

3

CONFIRMATION BIAS

Confirmation Bias is when analysts favour information that confirms their existing beliefs, skewing data analysis.

How to Identify: Look for patterns of selecting or interpreting data that confirm pre-existing beliefs or hypotheses.

How to Avoid: Actively seek out contradictory evidence and consider alternative explanations.

Example: A researcher focuses on patient testimonials supporting a new drug, ignoring contradictory evidence.

Aagam Deolasi

4

SELECTION BIAS

Selection Bias arises when the data sample does not represent the population being studied, leading to inaccurate conclusions.

How to Identify: Examine whether the data sample represents the population being studied.

How to Avoid: Use random sampling techniques and ensure unbiased data collection methods.

Example: Conducting a survey about smartphone usage exclusively among tech-savvy individuals, resulting in skewed data that does not reflect the broader population.

Aagam Deolasi

5

INTERPRETATION BIAS

Interpretation Bias occurs when analysts misinterpret data or draw conclusions that do not accurately reflect the underlying information.

How to Identify: Scrutinize interpretations of data to see if they align with the question being asked.

How to Avoid: Clearly define analysis objectives and validate interpretations with multiple perspectives.

Example: Interpreting a high bounce rate on a webpage as a sign of poor performance, without considering its simplicity or single-purpose nature.

Aagam Deolasi

6

INFORMATION BIAS

Information Bias results from incomplete or biased data sources that do not provide a comprehensive view of the phenomenon being studied.

How to Identify: Assess whether data sources provide a comprehensive view of the phenomenon being studied.

How to Avoid: Diversify data sources and consider potential biases inherent in each source.

Example: Using data from an online survey to analyze consumer preferences, which may not capture the opinions of individuals without internet access.

Aagam Deolasi

7

PREDICTIVE BIAS

Predictive Bias occurs when historical data used for predictions does not accurately reflect current or future conditions, leading to inaccurate forecasts.

How to Identify: Evaluate whether historical data accurately reflects current or future conditions.

How to Avoid: Regularly update models with current data and test predictions against real-world outcomes.

Example: Predicting customer demand based on pre-pandemic sales data, which fails to account for shifts in consumer behaviour during the COVID-19 pandemic.

Aagam Deolasi

8

IMPORTANCE of Data Privacy Regulations

1. Data privacy regulations are crucial as technology advances and data gains value.
2. 71% of countries have data security laws, with 9% drafting legislation, and **15% lacking** any.
3. The **GDPR**, adopted by the EU, governs personal data processing and applies universally.
4. **Personally identifiable information (PII)** includes data that identifies individuals, emphasizing the need for protection.

Aagam Deolasi

9

Data Security: The CIA Triad

Confidentiality, Integrity and Availability, known as the CIA Triad, is a summary guideline for organizational data security. **Confidentiality** ensures the privacy of data by restricting access through authentication and encryption. **Integrity** assures that the information is accurate and trustworthy. **Availability** ensures that the information is accessible only to authorized people.

Aagam Deolasi

10

The CIA Triad: CONFIDENTIALITY

1. Also known as **privacy**, it ensures that only authorized personnel access sensitive data.

2. Data is compartmentalized based on its sensitivity level, and employees receive training on safeguarding practices.

3. Methods include encryption, authentication, multi-factor authentication, and minimizing exposure.

Aagam Deolasi

11

The CIA Triad:
INTEGRITY

1. Ensures data remains accurate and trustworthy throughout its lifecycle.

2. File permissions, user access control, and version control prevent unauthorized access and accidental changes.

3. Hashing, a mathematical algorithm, verifies data integrity during transmission.

Aagam Deolasi

12

The CIA Triad:
AVAILABILITY

1. Focuses on maintaining equipment, performing repairs, and keeping systems up-to-date.

2. Backups and disaster recovery plans ensure data remains accessible during disasters.

3. Security measures like firewalls guard against cybersecurity threats to prevent downtime.

Aagam Deolasi

THANK YOU!!! FOR YOUR SUPPORT! For now...

Keep Learning, Keep Sharing, & Keep Following ***Aagam Deolasi***.

Aagam Deolasi