DATA SCIENCE_BWF_TASK_01

TOPIC-01: DEFINING DATA SCIENCE

DATA

Data surrounds us in everyday life, from the text we read to the phone numbers in our smartphones. Data is information, often in the form of facts or statistics.

DATA SCIENCE

The main goal of data science is to extract knowledge or to understand data (data can be both structured and unstructured). It uses scientific methods like statistics and probability. To gain practical insights gathered knowledge can be applied.

Data science itself is also a broad field touching many other disciplines like

- Databases
- Big Data
- Machine Learning (ML)
- Artificial Intelligence (AI)
- Visualization
- Infographics
- Human-Computer Interaction

WHERE TO GET DATA

Structured	Semi-Structured	Unstructured
Internet of Things (IoT)	Social Networks	Images or Videos
Behavioral Analysis	Group Dynamics	Texts
Surveys		Logs

WHAT WE CAN DO WITH THE DATA

It includes following steps

1. Data Acquisition: Process of collecting data by using different software tools, sensors and instruments etc.
2. Data Storage: After collecting data, second step is to store data. Data storing can be difficult if we have big data. There are several methods including
    - Relational Database
    - NoSQL Database
    - Data Lake

3. Data Processing: Converting original data into the form which can be used for visualization and model training.
4. Data Visualization: Graphical representation of data. Different charts, graphs and data visualization tools can be used to visualize the data.
5. Training a predictive model: Different Machine Learning techniques can be used to build a predictive model.

DIGITALIZATION AND DIGITAL TRANSFORMATION

Digitalization: Translate business processes into digital form.

Digital transformation: Applying data science techniques to this data to guide decisions can lead to significant increases in productivity (or even business pivot).

TOPIC-02: DATA SCIENCE ETHICS

By 2022, market trends tell us that 1-in-3 large organizations will buy and sell their data through online Marketplaces and Exchanges. Trends also indicate that by 2025, we will create and consume over 180 zettabytes of data.

The word "ethics" comes from the Greek word **"ethikos"** (and its root "ethos") meaning character or moral nature.

Data Ethics: Data ethics are now *necessary guardrails* for data science and engineering, helping us minimize potential harms and unintended consequences from our data-driven actions. It is essential for organizations to ensure that they handle data ethically to maintain trust with their customers and avoid legal issues.

ETHICS CONCEPT

It includes

1. **ETHICS PRINCIPLES:** The "shared values" that describe acceptable behaviors, and guide compliant actions, in our data & AI projects. It includes

- Accountability

- Transparency
- Fairness
- Reliability & Safety
- Privacy & Security
- Inclusiveness

2.ETHICS CHALLENGES

It can relate to the data privacy, ownership, informed consent *and* intellectual property rights for users.

Ethical challenges can arise from

- Dataset bias
- Data quality *i*ssues
- Unfairness
- Misrepresentation in algorithms

3. APPLIED ETHICS

1. **Professional Codes**
   Professional Codes are guidelines for ethical behavior in organizations, encouraging members to follow their principles through voluntary compliance, rewards, and penalties.
2. **Ethics Checklists**
   Ethics Checklists turn ethical principles into actionable tasks, helping practitioners ensure their projects meet ethical standards in a practical and trackable way.
3. **Ethics Regulations**
   Ethics Regulations involve complying with laws and standards that protect data privacy and promote ethical practices within organizations, like GDPR and HIPAA.
4. **Ethics Culture**
   Ethics Culture focuses on creating a shared commitment to ethical practices within organizations, encouraging early identification and resolution of ethical issues.

TOPIC-03: DEFINIG DATA

DATA

Data surrounds us in everyday life, from the text we read to the phone numbers in our smartphones. Data is information, often in the form of facts or statistics.

HOW DATA IS DESCRIBED

1. RAW DATA

Data that has come from its source in its initial state and has not been analyzed or organized.

2. QUANTITATIVE DATA

Quantitative data is numerical observations within a dataset.

It can typically be analyzed, measured and used mathematically.

EXAMPLES

- A country's population
- A person's height

3. QUALITATIVE DATA (Categorical Data)

Data that cannot be measured objectively like observations of quantitative data.

EXAMPLES

- Video Comments
- Favorite Color

4. STRUCTURED DATA

Data that can be in organized form i-e; in rows and columns.

Where Columns represent a value of a particular type, and rows contain the actual values.

EXAMPLES

- Spreadsheets
- Relational databases
- Phone numbers
- Bank statements

5. UNSTRUCTURED DATA

Unstructured data typically cannot be categorized into rows or columns and doesn't contain a format or set of rules to follow.

EXAMPLES

- Text files & messages
- Video files

6. SEMI STRUCTURED DATA

It is combination of both structured and unstructured.

EXAMPLES

- HTML
- CSV files
- JavaScript Object Notation (JSON)

SOURCES OF DATA

It is the starting point from where data is originated.

Primary data: Data generated by its users

Secondary Data: Data collected from a source that has collected data for general use.

Common Sources

- Databases
- Files
- Internet Sources
- API (Application Programming Interfaces)