Module 1: Introduction to Data Science

What is Data Science?

- Definition: Data Science is the interdisciplinary field of using scientific methods, algorithms, and systems to extract knowledge and insights from structured or unstructured data.
- Real-time Example: Retail companies analyzing customer behavior to predict future buying patterns.

What does Data Science involve?

- Steps: Defining objectives, data collection, data cleaning, exploratory data analysis, modeling, and deployment.
- Real-time Example: Healthcare analysts use data science to improve patient care by predicting illness based on health data.

Era of Data Science

- Development: From manual statistical analysis to automated and predictive analytics powered by machine learning.
- Real-time Example: Financial services use data science to detect fraudulent transactions in real-time.

Business Intelligence vs Data Science

- Comparison: Business Intelligence provides descriptive analytics looking at past data; Data Science uses machine learning to predict future trends.
- Real-time Example: E-commerce platforms use BI to report on past sales performance and data science to predict future inventory needs.

Life Cycle of Data Science

- Phases: Problem understanding, data preparation, modeling, evaluation, deployment.
- Real-time Example: Marketing agencies use the data science lifecycle to optimize ad campaigns based on user engagement data.

Tools of Data Science

- Python, R, SQL, Tableau, and machine learning frameworks like TensorFlow and PyTorch.
- Real-time Example: Data scientists use Python to process and analyze data for complex data visualization in market trend analysis.