Assignment 01

Fundamentals of Data Science

Instructors: Sir Irfan Malik , Sir Sheraz

Question 1: The Big Picture of Data Science

Answer:

1. Data Science as an Interdisciplinary Field:

- **Statistics:** Helps to analyze data and find patterns.
- **Computer Science:** Processes big data and builds models.
- **Domain Knowledge:** Understands the subject area for useful insights.

All three areas are important to solve real-world problems using data.

2. Difference Between Data Science and Machine Learning:

- **Data Science:** Involves collecting, cleaning, and analyzing data.
- **Machine Learning:** Uses data to train models for predictions.

Example:

- *Data Science:* Collecting and analyzing weather data.
- *Machine Learning:* Predicting future climate trends.

3. Importance of Soft Skills:

- Communication helps explain complex findings.
- **Example:** If results are not explained well, managers may misunderstand and make wrong decisions, leading to project failure.

Question 2: The Data Science Process in Action

Answer:

1. Key Stages of Data Science Process:

- **Problem Definition:** Understand the goal (recommend books).
- **Data Collection:** Gather user and book data.
- **Data Cleaning:** Fix errors and remove missing values.
- **EDA (Exploratory Data Analysis):** Find patterns in data.
- **Feature Engineering:** Create new useful features.
- **Model Building:** Train a recommendation model.
- **Model Evaluation:** Check model accuracy.
- **Deployment:** Use the model in an app.
- **Monitoring:** Track and improve performance.

2. Importance of EDA:

- **Detecting Outliers:** Remove unusual data points.
- **Checking Data Types:** Make sure data is in correct format.

3. Model Evaluation Metric:

- **Precision:** Measures how many recommended books users liked.

Question 3: Understanding Data Attributes

Answer:

1. Classification of Attributes:

Attribute	Type	Explanation
Height (in cm)	Numerical	Measurement data, continuous.

Favorite Subject	Nominal	Categories without order.
Exam Pass/Fail	Binary	Two options: Yes or No.
Student ID	Nominal	Unique identifier, not numerical.

2. Asymmetric Binary Attribute:

- **Exam Pass/Fail**
- **Explanation:** Misclassifying "Fail" as "Pass" means missing support for the student.

3. Why Not Average Student ID:

- **Reason:** Student ID is nominal.
- **Explanation:** IDs are labels for identification, not for calculation.