Xian Hui B. Cheng
BSIT 3-5
Cheng

1. What is data science?

It is a field that focuses on extracting insights & knowledge from data. It combines technique from statistics, mathematics, and more to analyze large and complex data sets.

Goal: Find patterns, make predictions, and drive informed decisions based on data. It's used in various industries to solve problems, optimize processes and understand trends through data-driven approaches.

Real Life Application

Example: Data Science is used not only on the tech field, but also in healthcare. It is used to predict disease outbreaks, improve diagnostic accuracy, & personalized treatment plans. Hospitals also use predictive models to manage patients.

2. Differentiate between Business Intelligence (BI) and Data Science?

Business Intelligence (BI)

→ use of tools, technologies, and processes to analyze historical data and generate insights that support decision-making. It focuses on descriptive analytics — reporting & summarizing past business performance to improve decision making.

Data Science

→ extracts insights from structure & unstructure data.

Main Differences

Aspect	Business Intelligence	Data Science
Focus	Analyze past data to describe what happened.	uses data to predict future outcomes & prescribe actions
Purpose	supports decision making through reports & dashboards based on historical data.	solves complex problems & makes predictions using algorithms & models
Technique Used	Reporting, querying, data aggregation	Machine learning, statistical analysis, data mining
Time Orientation	Retrospective (focuses on past data)	Prospective (focuses on future trends & predictions)

3. Discuss the role of a ~~visual~~ data visualization in data science. How can effective visualization improve the interpretation of data & findings

→ Data visualization is important in data science, it serves as the one who converts complexitivity data sets into easily digestable visual formats like charts, graphs, and dashboards. This helps people get key insights in just one glance

[Importance]

    1. Simplification of Complex Data
→ breaks it down in just one glance & makes it easier to interpret

    2. Highlight patterns & trends
→ good visualization shows trends that are not seen/obvious in raw data

    3. Enhancing communication
→ makes it easier for people to comprehend what the data is saying and make decision based on insights

4. ~~Draw~~ Give atleast (3) Fundamental Concerts of Data science & explain.

1. Exploratory Data Analysis
→ involves using visual & statistical techniques to explore data & uncover patterns, trends & relationships. It helps data scientists gain insights into the structure of the data & understand key features before applying advanced models

[Importance] Detect outliers, identify missing values & guide future analysis

2. Statistical Inference
→ using data from a sample to draw conclusions about a larger population. Includes estimating population parameters, hypothesis testing and determining the statistical significance of the findings

[Importance] Essential for making predictions based on limited data

3. Machine learning
→ enables computers to learn from data & make predictions/classifications. involves training models, selecting algorithms, & tuning hyper parameters

Importance

Allows data scientists to build models that automate decision & solve problems. It allows building of algorithms that maker accurate prediction/ classify based on patterns.

Common techniques used:

① supervised learning
→ model is trained on labeled dataset
→ your input data are paired w/ the correct output
          ↓   ↓
       Features   label
→ Goal: learn a mapping from inputs to outputs so model can correctly the corret label for new unseen data.

② Unsupervised learning.
→ model is trained w/out explicit labels
→ Goal: Identify patterns, relationships, structures whitin the data.

③ Reinforcement learning
→ Agent learns by interacting w/ its enviromrent & receiving feedback in the form of rewards/penalties.
→ Used in times where sequences of decijions needs to be made.