Aufera Data Science

Lecture: 01

Data Science:

It's an interdisciplinary field that includes:

1. statistics
2. AI
3. Programming
4. Domain expertise

Machine Learning Vs Data science :

Machine Learning	Data science
* Focus on studying models and Algorithms that learn from data	* works one step beyond ML solves real-world problems. Follows an end to end process from problem formulation to presentation

Data Science End-to-End process:

1. Problem formulation (Defining the Question)
2. Data acquisition (Gathering relevant data)
3. Data preparation (cleaning and transforming)
4. Model development (Building predictive models)
5. Model deployment (Implementing in production)
6. Insight communication (sharing results)

Skill set Required for Data science:

- Statistics (understanding data distribution)
- Machine Learning (Building predictive models)
- Python programming (Implementing solutions)
- Data visualization (creating insightful graphs)
- communication skills

Lecture: 02

Step 1: Problem Formulation

- domain expertise helps in formulating a problem
- Identifying input data
- selecting the appropriate Algorithm
- Define the output (predicted price)
- Documenting the problem statement

Step 2: Data Acquisition

- Business-owned data repositories
- kaggle Dataset
- Government Datasets

step 3: Data preparation

- understanding your data (EDA)

step 4: Data Preprocessing

- Handling missing values
- Removing inconsistencies
- Data transformation

→ Data reduction

step 5 : Data Analysis

→ On this stage models are developed

→ Model can be ① classification ② Regression ③ clustering ④ statistical Modeling etc] train the model

→ Evaluate the Models

→ Deploy Top models on server

step 6 : communication & visualization

→ Data driven insights must be presented effectively.

[A picture is worth 1000 words]