Data Science Handbook – Page 1

Data science is an interdisciplinary field that uses scientific methods, algorithms, and systems to extract insights from structured and unstructured data. It combines skills from statistics, computer science, and domain knowledge to make data■driven decisions. A data scientist typically works with large datasets, performs data cleaning, explores patterns, builds machine learning models, and evaluates results.

The data science workflow includes data collection, data preprocessing, exploratory data analysis (EDA), model building, and model deployment. Tools like Python, R, SQL, and visualization libraries help convert raw data into meaningful insights.

Data Science Handbook – Page 2

Key components of data science include:

1. **Data Collection:** Gathering data from databases, APIs, logs, or files.
2. **Data Cleaning:** Handling missing values, duplicates, and errors.
3. **Exploratory Data Analysis:** Understanding data shape, summary statistics, and patterns.
4. **Machine Learning:** Using algorithms like regression, decision trees, clustering, and neural networks.
5. **Model Evaluation:** Measuring accuracy with metrics like precision, recall, and RMSE.

Data science enables businesses to make better decisions, improve customer experience, optimize operations, and develop intelligent applications.