Introduction to Data Science

Welcome to the world of data science! This presentation will guide you through the basics of data science and its applications in various fields.

What is Data Science?

Data science is a multidisciplinary field that involves extracting knowledge and insights from data. It combines computer science, statistics, mathematics, and domain expertise to solve real-world problems.

Data Collection

Gathering raw data from various sources.

Data Cleaning

Dealing with missing values, outliers, and inconsistent data.

Data Analysis

Exploring, transforming, and visualizing data to uncover patterns.

Data Modeling

Building predictive models to forecast outcomes and make decisions.

The Data Science Process

The data science process is iterative, involving several steps to extract meaningful insights from data.

1. **Define the Problem**

   Clearly define the business problem you want to solve.

2. **Data Acquisition**

   Collect relevant data from various sources.

3. **Data Exploration**

   Analyze data to understand its characteristics.

4. **Model Building**

   Develop predictive models to solve the problem.

5. **Model Evaluation**

   Assess the performance of the model.

6. **Deployment**

   Implement the model to make predictions.

Data Collection and Preprocessing

Data collection involves gathering raw data from different sources, such as databases, APIs, and web scraping.

Data Sources

Databases, APIs, social media, web scraping.

Data Cleaning

Handling missing values, outliers, and inconsistent data.

Data Transformation

Converting data into a format suitable for analysis.

Exploratory Data Analysis

Exploratory data analysis (EDA) involves using statistical and visual techniques to understand data characteristics and discover patterns.

1

Descriptive Statistics

Summarizing data using measures like mean, median, and standard deviation.

2

Data Visualization

Creating charts and graphs to visualize data trends and relationships.

3

Hypothesis Testing

Evaluating hypotheses about the data using statistical tests.

Machine Learning Fundamentals

Machine learning enables computers to learn from data without explicit programming, enabling them to make predictions and decisions.

Data Input

Provide the model with training data.

Model Training

The model learns patterns from the data.

Model Evaluation

Assess the model's performance on unseen data.

Model Deployment

Use the trained model to make predictions.

Supervised and Unsupervised Learning

Machine learning algorithms can be categorized into supervised and unsupervised learning, based on the type of data and task.

Type	Description	Example
Supervised	Training with labeled data to make predictions.	Predicting house prices based on features.
Unsupervised	Discovering patterns in unlabeled data.	Clustering customers into different groups.

Applications of Data Science

Data science has revolutionized various industries by providing insights and solutions to complex problems.

Healthcare

Disease prediction, personalized medicine, drug discovery.

Finance

Fraud detection, risk assessment, investment strategies.

Marketing

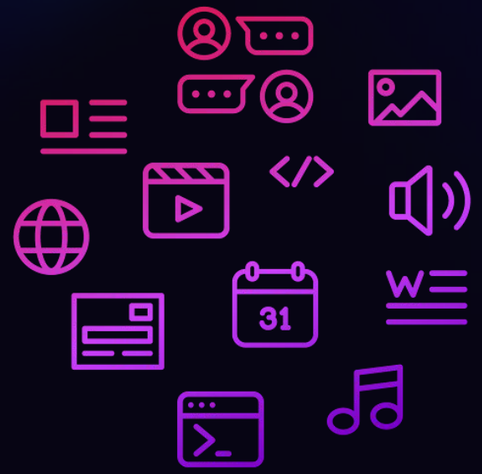
Targeted advertising, customer segmentation, campaign optimization.

E-commerce

Personalized recommendations, inventory management, price optimization.

Structured & Unstructured data

Structured data refers to information that is organized in a specific format, such as tables or databases. On the other hand, unstructured data refers to data that does not have a predefined structure, like text documents or social media posts.



UNSTRUCTURED DATA

STRUCTURED DATA