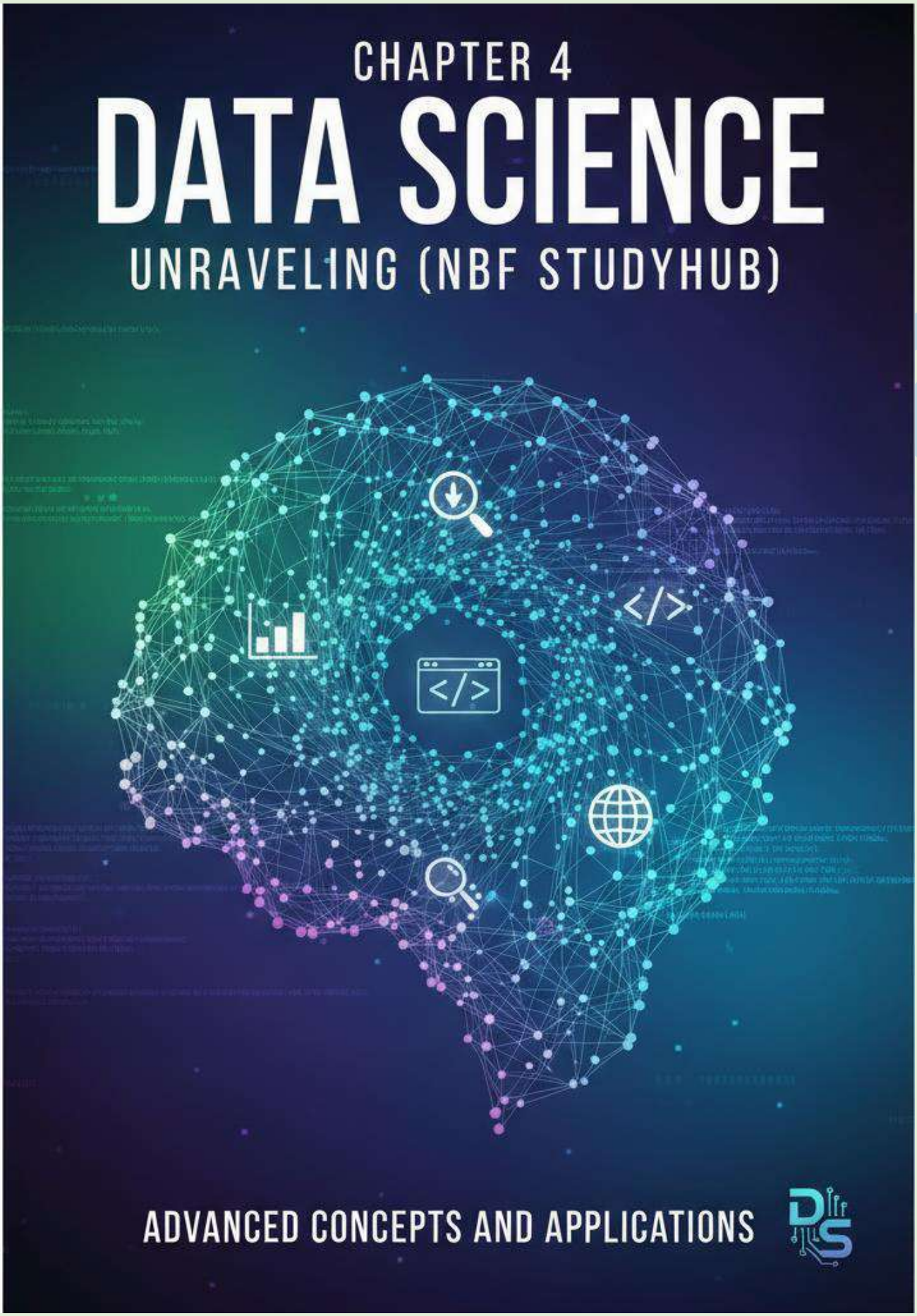

CHAPTER 4
DATA SCIENCE
UNRAVELING (NBF STUDYHUB)
ADVANCED CONCEPTS AND APPLICATIONS

Chapter: 04

Notes on Data Science

Simplest & Easy for Exam Preparation

Data Science and Its Scope

Data Science is an interdisciplinary field that focuses on collecting, cleaning, analyzing, and interpreting large volumes of data to extract meaningful insights. It combines knowledge from statistics, computer science, and machine learning to help organizations make informed decisions.

The scope of Data Science is vast and continues to grow as more industries realize the value of data-driven strategies. Here are some key areas where Data Science plays a crucial role:

- **Business & Marketing:** Helps companies understand customer behavior, segment markets, improve product targeting, and increase sales by analyzing consumer data and market trends.
- **Healthcare:** Enables prediction of diseases, personalized treatment plans, and efficient management of patient data to improve healthcare outcomes.
- **Finance:** Used for fraud detection, credit scoring, risk assessment, algorithmic trading, and financial forecasting to make better investment decisions.
- **E-commerce:** Supports personalized recommendations, inventory management, and customer experience enhancement by analyzing shopping patterns and preferences.
- **Government:** Assists in managing traffic, predicting criminal activities, improving public safety, and optimizing public services through data analysis.
- **Education:** Facilitates monitoring student performance, identifying learning gaps, and designing personalized education programs.
- **Social Media & Entertainment:** Powers content recommendation systems, sentiment analysis, and user engagement tracking to enhance user experience.

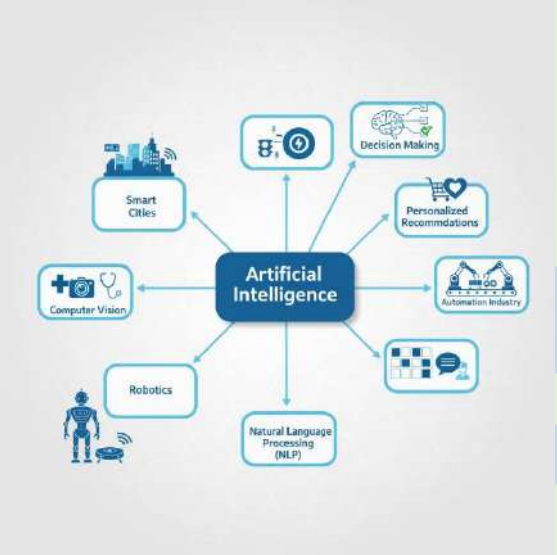
With the increasing amount of data generated daily, the demand for skilled data scientists and analysts is growing rapidly. Data Science not only helps organizations solve complex problems but also uncovers new opportunities for innovation and growth.

4.1.2 Artificial Intelligence and Its Scope

Artificial Intelligence (AI) means the **ability of a machine to think and act like humans**. It can solve problems, understand human language, and interact with the environment in smart ways. The idea is not new — in 1950, British mathematician **Alan Turing** introduced the *Turing Test*, which checks if a machine can act so intelligently that humans cannot tell it apart from another human. Today, robots that can pass this test are considered very smart.

Main Areas of AI Scope (in simple words):

1. Decision Making:
AI can help take the *best decision* by using data. For example, it can analyze many possibilities and choose the most efficient one.



2. Personalized Recommendations:
It can give suggestions based on what the customer liked or did before. For example, YouTube recommending videos that match your viewing history.

3. Automation Industry:
AI can do *boring or repetitive tasks*. This is useful in factories (like car manufacturing) or tasks such as image and video analysis. Smart devices and the Internet of Things (IoT) also work with AI to make everyday life easier.

4. Natural Language Processing (NLP):
AI helps machines understand and respond to human language. Examples include ChatGPT, chatbots, or voice-activated devices. You can even turn on an air conditioner just by saying "AC on".

5. Robotics:
Modern robots can talk, cook, clean, and do many human tasks — all because of AI.

6. Healthcare:
AI can give advice for *personalized treatments*. It can check medical images and help in diagnosis.

7. Computer Vision:
A special field of AI where computers learn to understand images and videos. For example, detecting objects in photos.

8. Smart Cities:
AI helps design **efficient cities** with better energy use, traffic control, and public services.

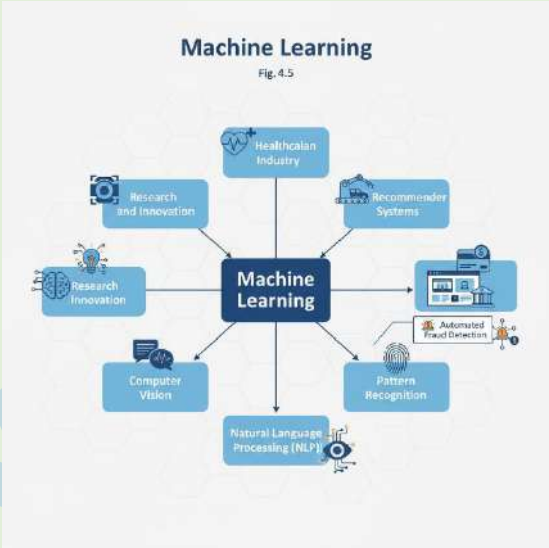
9. AI Agents:
Smart assistants like Siri, Alexa, Google Assistant, Cortana, and ChatGPT work using AI.

4.1.3 Machine Learning and Its Scope

Machine learning is a branch of artificial intelligence where computers learn from data and improve their performance without being directly programmed. It is used in many areas of our daily life and helps machines make decisions by finding patterns in data.

Scope of Machine Learning

- **Healthcare:**
  Machine learning helps doctors predict diseases, suggest treatments, and analyze medical images for better diagnosis.
- **Automation Industry:**
  Machines and robots can perform tasks automatically, like assembling cars or operating automatic doors and lights, without human help.
- **Recommender Systems:**
  Websites like YouTube, Netflix, and Amazon use machine learning to suggest videos, movies, or products based on what you have watched or bought before.
- **Finance and Banking:**
  Machine learning is used to detect fraud by finding unusual activities, such as sudden large transactions or strange network traffic.
- **Pattern Recognition:**
  It helps computers recognize patterns in data, such as handwriting, speech, or images.
- **Natural Language Processing (NLP):**
  Machine learning allows computers to understand and respond to human language. Virtual assistants like Siri, Google Assistant, Alexa, and Cortana use NLP to follow voice commands.
- **Computer Vision:**
  This field helps computers understand and analyze images and videos. Examples include face recognition in smartphones and using Google Lens to identify objects.
- **Research and Innovation:**
  Machine learning is used in scientific research to analyze data, make predictions, and discover new patterns.

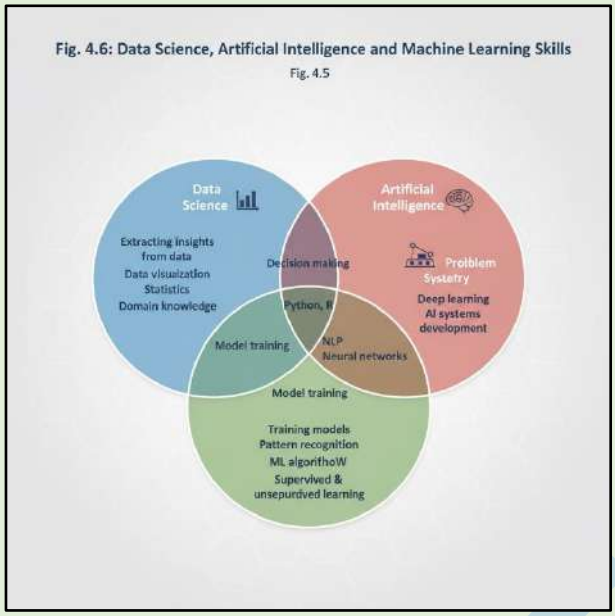


Key Point

The main power behind data science and artificial intelligence comes from machine learning algorithms. These algorithms help computers learn from experience and get better over time, just like humans do.

4.1.4 Data Science, Artificial Intelligence, and Machine Learning Skills

To work in Data Science, Artificial Intelligence (AI), and Machine Learning (ML), you need some important skills. These skills help you build smart systems that can learn, make decisions, and find useful information from data.



Main Skills Needed

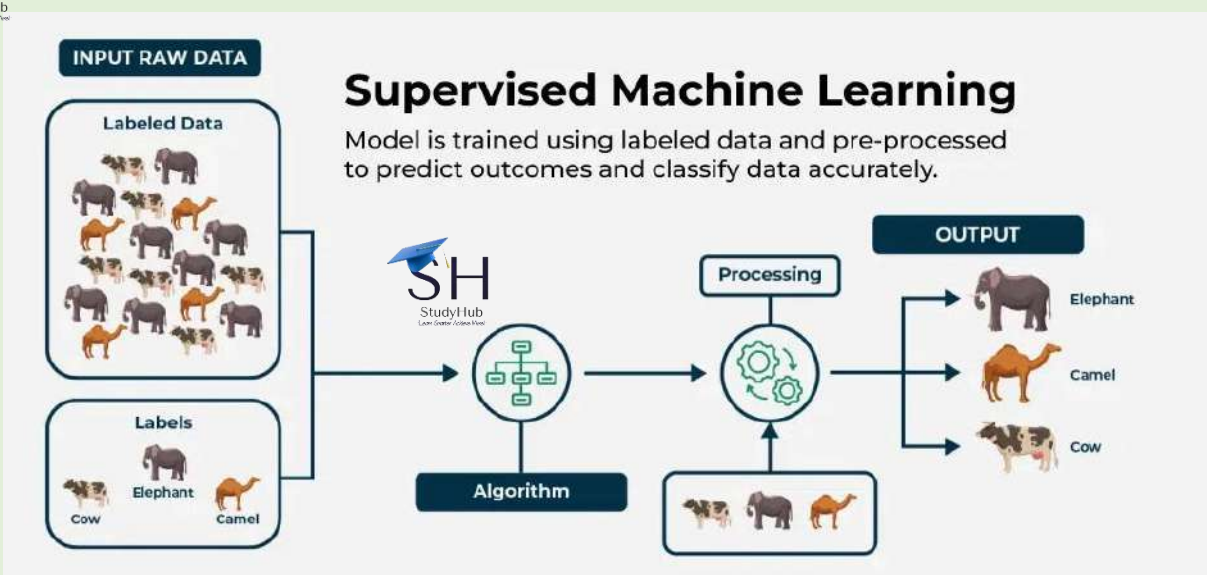
- **Programming Languages:**
You must know how to code. The most popular languages for AI and ML are Python and R because they are easy to learn and have many useful libraries.
- **Machine Learning:**
You should understand how machines can learn from data. This includes knowing about different algorithms and how to use them.
- **Deep Learning and Neural Networks:**
These are advanced parts of machine learning. Deep learning helps computers solve complex problems, like recognizing faces or understanding speech.
- **Natural Language Processing (NLP):**
This skill helps computers understand and work with human language, like chatbots or voice assistants.
- **Frameworks and Tools:**
Tools like TensorFlow are used to build and train machine learning models. Knowing how to use these tools is very important.
- **Domain Knowledge:**
You should also know about the specific field you are working in, like healthcare, finance, or marketing. This helps you solve real-world problems better.
- **Data Science Skills:**
You need to know how to collect, clean, and analyze data to find useful patterns and make good decisions.

4.1.5 Machine Learning Models

There are three main types of machine learning models. Each model helps computers learn in a different way.

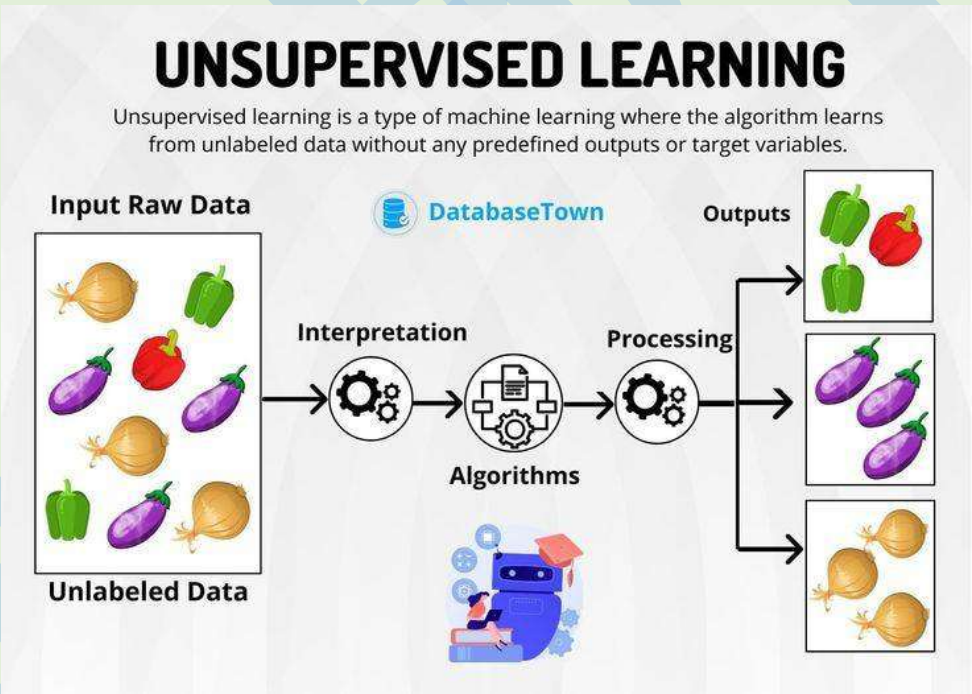
Supervised Machine Learning Model

In supervised learning, the computer learns from labeled data. This means the data already has answers, and the computer tries to learn the pattern to predict future answers. For example, if you show a computer pictures of cats and dogs with their names, it learns to tell the difference between them.



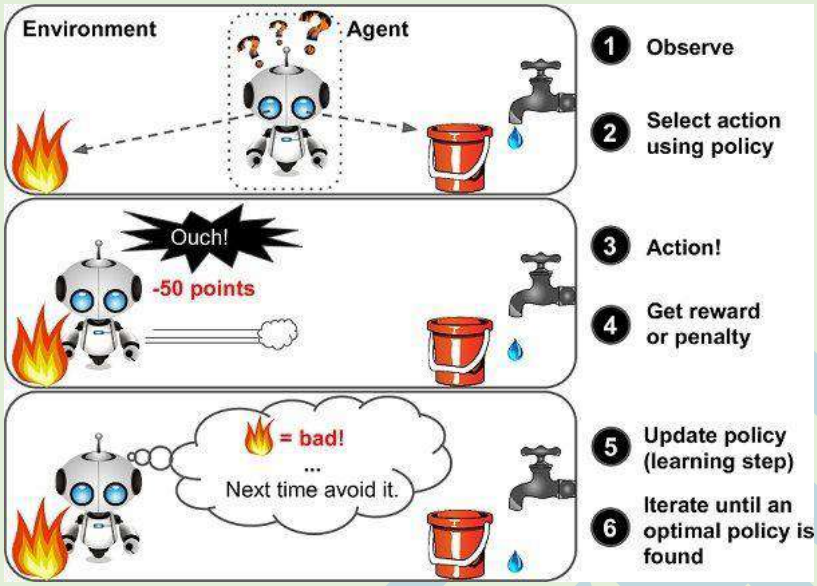
Unsupervised Machine Learning Model

In unsupervised learning, the computer gets data without any labels or answers. It tries to find patterns or groups in the data by itself. For example, if you give a computer many pictures without telling what they are, it will try to group similar pictures together.



Reinforcement Machine Learning Model

In reinforcement learning, the computer learns by trying different actions and getting rewards or punishments. It learns from its mistakes and successes, just like humans do. For example, a robot learns to walk by getting rewards for moving forward and learning from falling down.



Do You Know?

- **TensorFlow:**
  TensorFlow is a tool that helps computers learn from data. It works like a robot learning from experience, just like humans do.
- **Deep Learning:**
  Deep learning is a special way for computers to learn using layers of artificial neural networks. It helps computers solve complex problems, like recognizing faces or understanding speech.
- **Neural Network:**
  A neural network is a system made to work like the human brain. It is built from artificial neurons that make decisions, similar to how our brain works.

4.1.6 Supervised Learning Model

Supervised learning is a way to train a computer using labeled data, where each example has inputs and the correct answer. The computer studies many such examples to learn patterns, and later predicts answers for new, unseen inputs.

How it works

- You give the machine many examples with answers. For example, pictures of fruits with their names.
- Each example has features, like shape, color, and texture.
- The machine learns the relationship between features and the correct name.
- After training, you give new features without the answer, and the model predicts the most likely name.

Benefits and limits

- Benefits: High accuracy when labels are correct; clear goal to learn.
- Limits: Needs lots of labeled data; can make mistakes if training data is biased or poor.

4.1.7 Unsupervised Learning Model

Unsupervised learning is a type of machine learning where the computer is given data without any labels or correct answers. The computer looks for patterns, similarities, and differences in the data and groups similar items together.

How it works

- The machine receives a lot of data, but there are no answers or labels.
- It tries to find hidden patterns or groups in the data by itself.
- After learning, it can place new data into the correct group based on what it has learned.

Simple example

- You give the computer many news articles without telling what type they are.
- The computer studies the words and topics in each article.
- It groups the articles into categories like national politics, international politics, entertainment, medical, and technology.
- When you give a new article, like "Scientists successfully saved the cancer patients at last stage," the computer puts it in the medical news group.

Benefits and limits

- Benefits: Useful when you don’t have labeled data; helps discover hidden patterns.
- Limits: The groups may not always make sense to humans; results can be harder to understand.

4.1.8 Reinforcement Learning Model

Reinforcement learning is a type of machine learning where a computer (called an agent) learns by getting feedback from its actions. The agent tries different actions in an environment and receives rewards for good actions or penalties for bad actions. Over time, the agent learns to make better decisions to get the highest rewards.

Key Terms

- **Agent:** The machine or computer that learns and makes decisions.
- **Environment:** The place or situation where the agent acts (for example, a chessboard).
- **Actions:** The moves or steps taken by the agent.
- **Rewards:** Points or positive feedback for helpful actions.
- **Penalties:** Points taken away or negative feedback for unhelpful actions.

How it works

- The agent tries different actions in the environment.
- If an action helps reach the goal, the agent gets a reward (like points).
- If an action is not helpful, the agent gets a penalty (loses points).

- The agent uses this feedback to learn which actions are best.
- This process continues until the agent learns the best way to act.

Simple example

- Imagine you are learning to play chess.
- If you start the game by moving the knight and you win, you learn that this is a good move.
- If you start by moving the queen and you lose, you learn to avoid this move.
- By winning and losing, you learn which moves are better.
- The computer agent does the same: it tries moves, gets feedback, and improves over time.

Common uses

- Training robots to walk or play games
- Self-driving cars learning to avoid obstacles
- Game AI learning to win against human players

Benefits and limits

- Benefits: Learns from experience; can solve complex problems by trial and error.
- Limits: Can take a long time to learn; needs many tries to get good results.

4.1.9 Choosing an Appropriate Machine Learning Model

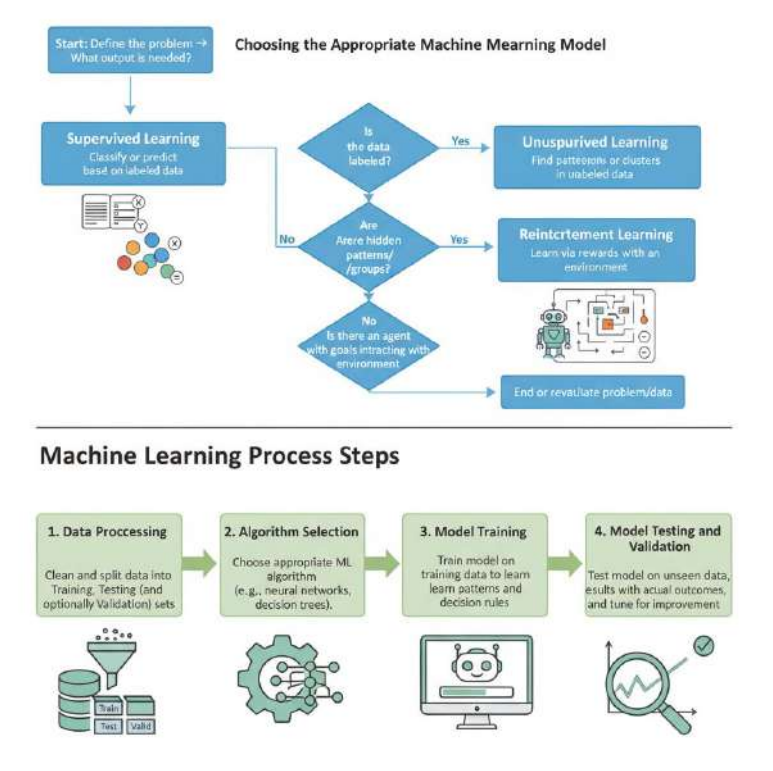
When you face a problem and want to use machine learning, you need to choose the right model. Here's how you can decide:

1. Define the Problem

- What do you want to achieve? Do you want to identify items, group them, or predict something for the future?

2. Analyze the Data

- **If your data has labels** (like questions with answers), use a **supervised learning model**. Example: Classifying emails as spam or not spam.
- **If your data has no labels** and you want to find groups or patterns, use an **unsupervised learning model**. Example: Grouping customers by shopping habits.
- **If your problem involves an agent learning by trial and error** to reach a goal, use a **reinforcement learning model**. Example: Training a robot to walk or play a game.



Machine Learning Process Steps

1. **Split the Data:**
   - Clean and prepare your data.
   - Divide it into **training data** (to teach the model) and **testing data** (to check the model). Sometimes, a small part is kept for **validation**.
2. **Choose an Algorithm:**
   - Pick a suitable method, like a neural network or decision tree, to train your model.
3. **Train the Model:**
   - Use the training data to help the model learn patterns and make decisions.
4. **Test and Validate:**
   - Give the model new data (testing and validation data) and see how well it predicts the answers.
   - If the results are not good, you can try a different algorithm or adjust the settings to improve performance.

Summary:

- First, understand your problem and data.
- Choose supervised, unsupervised, or reinforcement learning based on your needs.
- Follow the steps: split data, choose algorithm, train, and test the model.

4.1.10 Artificial Intelligence, Machine Learning, and Data Science: Similarities and Differences

Artificial Intelligence (AI) means making computers or machines smart so they can do tasks that usually need human intelligence.

Machine Learning (ML) is part of AI. It teaches machines to learn from data by themselves without being told every step.

Data Science is about studying data to find useful information, which helps in making good decisions.

Similarities among AI, ML, and Data Science

- **Data-Driven:** All use data to get results.
- **Algorithms and Models:** They work with algorithms (step-by-step procedures). For example, AI uses search algorithms, ML uses learning algorithms, and Data Science uses statistics.
- **Automation Goal:** All aim to automate tasks that humans used to do manually.

Differences among AI, ML, and Data Science

Aspect	Artificial Intelligence	Machine Learning	Data Science
Tools and Techniques	Neural networks, expert systems, natural language processing	Supervised, unsupervised, reinforcement learning, deep learning	Statistical analysis, data mining, data visualization
Applications	Speech recognition, image recognition, chatbots, self-driving cars	Recommendation systems, fraud detection, spam filtering, marketing	Business intelligence, healthcare analytics, market analysis
Output	Systems that mimic human behavior	Predictions and classifications	Insights, reports, and visualizations

---

Difference Summary

Data Science	Artificial Intelligence	Machine Learning
Study of data to find insights	Imitates human behavior	A part of AI that learns from data
Focus on cleaning, analyzing data	Focus on decision making	Focus on training model to predict
Uses statistics and visualization tools	Uses ML algorithms to mimic humans	Uses algorithms to learn and predict

---

Churn Prediction

Churn prediction means finding out which customers might stop using a company's products or services.

- Used in business and marketing to keep customers.
- Example: A telecom company gets many complaints from some customers. Using churn prediction, they can guess these customers may leave.
- Keeping old customers is often better than finding new ones.

How Churn Prediction Works:

1. Collect customer data (feedback, buying frequency).
2. Use algorithms to analyze data.
3. Find customers who may leave.
4. Send special offers to keep those customers.

Behavioural Segmentation

Behavioural segmentation means grouping customers based on how they act or behave with a product or service.

- Helps businesses make better marketing plans.
- Groups customers using:
  - Purchase frequency (how often they buy)
  - Brand loyalty (how long they stay customers)
  - Usage occasions (when they use products)
  - Response to advertisements or offers

Example: A clothing store tracks how often customers buy formal clothes and gives special offers to those who might stop buying.

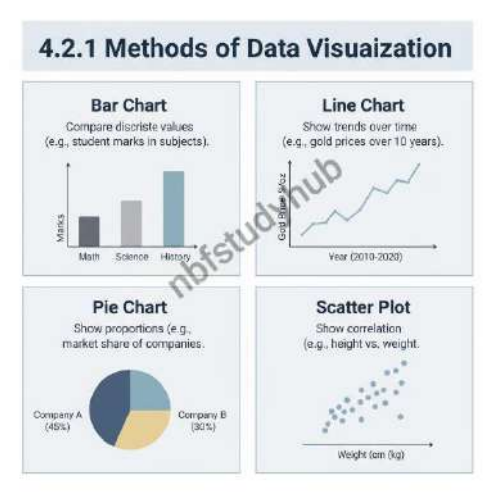
4.2 Data Visualization

Data Visualization means showing data or information in the form of pictures, like charts, graphs, and maps. This makes it easier to understand the data and helps us make better decisions in science, arts, business, and everyday life. Instead of looking at raw numbers, we use images to understand the story that data tells. Data visualization turns complex information into easy and useful insights.

By using charts and graphs, we can quickly see patterns or trends in data. Some methods even let us interact with the data to find the best answers. There are many ways to visualize data, and each helps us understand the information better to improve business, education, and services.

4.2.1 Methods of Data Visualization

The methods of data visualization are tools or techniques we use to prepare data for showing it clearly. Here are some common ones:

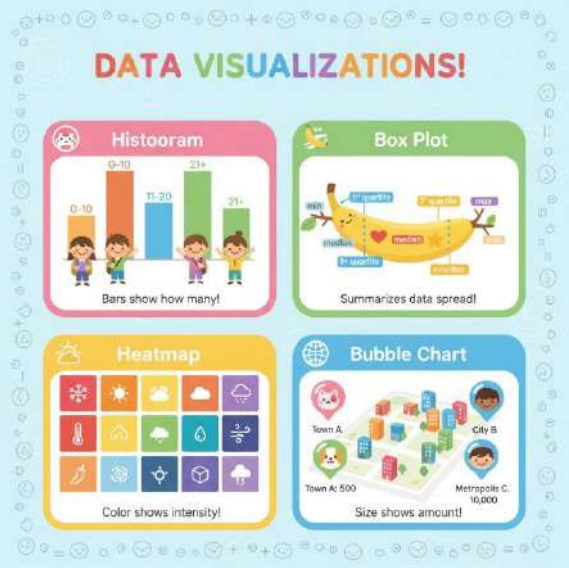


- **Bar Chart:**
  This uses bars to show comparison between different groups or categories. The length of the bar shows the size of the value.
  Example: Marks of students in different subjects or sales of different products.
- **Line Chart:**
  This shows how data changes over time. Points are connected by lines to show trends or patterns.
  Example: Prices of gold over the years or temperature changes during a year.
- **Pie Chart:**
  This shows how parts make up a whole. It looks like a circle divided into slices where each slice represents a part of the total.
  Example: Market shares of different companies, showing which company has the largest share.
- **Scatter Plot:**
  This shows the relationship between two sets of numbers using dots. It helps us see if two things are connected or not.
  Example: Relationship between height and weight of students, or head size and circumference.

Histogram

A histogram is a type of graph that helps us see how a single numerical value is spread out. It shows how often each range of numbers occurs. For example, if you want to see how many students scored in certain ranges in a class test (like 0-10, 11-20, and so on), a histogram will show this by drawing bars for each score range. The taller the bar, the more students scored in that range.

Box Plot (Box-and-Whisker Plot)



A box plot is a special graph that summarizes important parts of a set of data. It uses five main numbers: minimum (lowest value), first quartile (25% of data below this), median (middle value), third quartile (75% of data below this), and maximum (highest value). This helps you quickly understand how the data is spread and if there are any extreme values. For example, a company might use a box plot to show how salaries of its employees are distributed.

Heatmap

A heatmap is a colorful way to show data arranged in rows and columns (like a table or matrix). Each value in the table is shown with a color, and the color's darkness or brightness tells you how big or small the value is. For example, scientists use heatmaps to study weather patterns from space images or to see how climate has changed over time. Another example is using a heatmap to study different parts of a flower by color intensity.

Bubble Chart

A bubble chart is like a scatter plot but with an extra feature: the size of each point (bubble) changes depending on a third piece of information. So, instead of just showing where points lie on two axes, it also shows how big or important each point is based on bubble size. For example, you could show cities on a map where the position shows location and the bubble size shows population.

4.2.2 Types of Data Visualization

Data visualization means showing data or information in a picture or graph to make it easier to understand. Different kinds of data need different ways to show them. Here are some common types of data visualization:

Quantitative Visualization

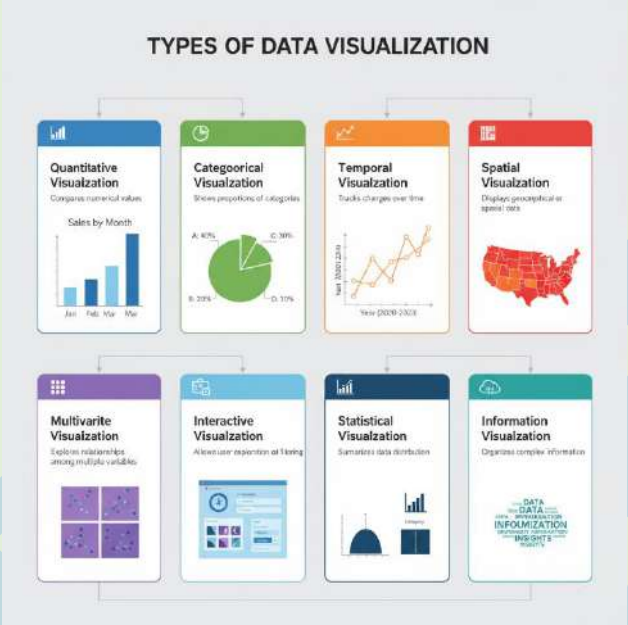
This is used to show numbers or amounts. It is useful when dealing with data you can count or measure. For example, a bar chart that shows the sales of a company month by month gives a clear picture of the numbers.

Categorical Visualization

This type shows data that fits into groups or categories. It helps to compare different parts of a whole. A good example is a pie chart that shows how much market share each company has in an industry.

Temporal Visualization

Temporal visualization shows how data changes over time. This is used for time-based data, like daily or yearly records. Line graphs are often used here, such as a line graph showing temperature changes over a week.



Spatial Visualization

This type shows data related to places or locations on maps. It helps us understand information connected to geography or space. For example, a heatmap showing which areas have more people living in them.

Multivariate Visualization

Multivariate visualization is for data that has many variables or factors. It shows how different things are related to each other using tools like scatter plots. For example, a graph that shows the link between people's income, age, and spending.

Interactive Visualization

Interactive visualization lets users click or choose different options to explore the data themselves. This happens on computers or tablets, often through dashboards. Dashboards help people filter and look at data in different ways to find useful information.

Statistical Visualization

This shows statistical facts like how data is spread out or how two things are connected. Common tools used are histograms, box plots, and scatter plots. They help understand things like averages and patterns.

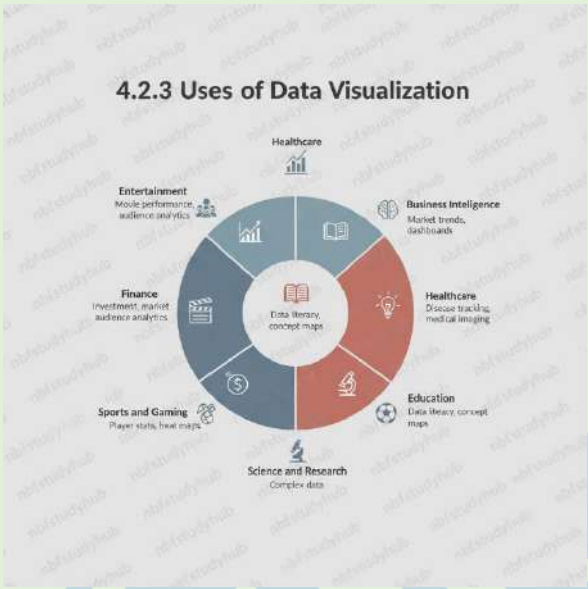
Information Visualization

This type is used for complex or abstract data which is hard to explain in simple numbers. It uses diagrams like network maps or word clouds. For example, a network diagram can show how friends are connected on social media.

4.2.3 Uses of Data Visualization

Like Artificial Intelligence, data science and machine learning, data visualization is useful in almost all the fields of life. Some of them are as follows:

Business Intelligence: Data visualization helps to make data driven, well informed decisions. It is used to find market trends and helps to track and improve performance.



Healthcare: It helps to visualize the impact of various diseases affecting the patient. It is helpful to track disease and visualize the spread of disease.

Education: Data visualization is very helpful to teach data literacy skills, concept building, creative thinking, and critical thinking.

Science and Research: It is useful to visualize complex findings, very huge and complex data such as complex scientific data received from satellite in the form of photographs.

Sports and gamming: It is useful to visualize performance of players whether they are playing football in a ground or chess players playing in online tournaments. It is also helpful in sports broadcasting, and other predictions.

Finance: It is helpful to analyze market trends, to track portfolio performance and to identify investment opportunities.

Entertainment: It helps the entertainment industry, to visualize movie performance data to predict future trends. It helps in content optimization by visualizing audiences' insights and trends.

4.2.4 Advantages / Benefits of Data Visualization

Data visualization means presenting data in the form of **charts, graphs, and pictures** instead of only numbers. This makes the data easy to understand even for people who are not technical. It helps in taking correct decisions faster and also shows patterns or trends clearly.

Main Benefits of Data Visualization:

- It helps **non-technical people** to understand data easily. For example, a shop owner can see a sales graph and quickly know if sales are increasing or decreasing.
- It helps decision-makers to take **better and informed decisions** based on accurate information.
- It improves the **performance of businesses and services** because they can see what is working and what needs improvement.
- It supports **better communication and teamwork** because everyone understands the situation through graphs and visuals.

- It **saves time** by showing results quickly instead of reading long reports.
- It allows **fast and correct decisions** in less time as compared to manual checking or calculations.

Extra Knowledge:

Today, companies use many tools for data visualization such as **Tableau, Microsoft Power BI, and Google Data Studio**. These tools convert raw data into meaningful visuals that can be understood easily.

4.2.3 Database and Machine Learning

A **database** is a structured way of storing data in **tables** which have rows and columns, just like spreadsheets. Databases are managed by special software called **DBMS (Database Management Systems)**. Common examples are **Oracle, MS Access, and MySQL**. Databases are very important because they keep data clean, organized, and easy to access.

A **Machine Learning (ML)** system is a technology in which computers learn from **datasets** and make decisions or predictions without being directly programmed. For example, Google Search suggestions, YouTube video recommendations, and Netflix movie suggestions all use machine learning.

Database Features:

- Stores data in tables in an organized way.
- Provides clean data by removing errors and duplicates.
- Data can be retrieved easily using **SQL queries**.
- Can handle both small and huge amounts of data (from 50 students' records to millions of records).
- Supports real-time storage and retrieval of data.
- Includes tools like SQL for data analysis.

Machine Learning Features:

- Needs data in the form of structured datasets (rows and columns).
- Works best when the data is clean and error-free.
- Uses algorithms to access and process data.
- Needs integrated and reliable data for accurate results.
- Can handle very large datasets (hundreds, thousands, or millions of records).
- Requires real-time data to update models and give correct outputs.
- Works with database tools to provide **fast and accurate solutions**.

Relationship Between Database and Machine Learning

Feature	Database (DB)	Machine Learning (ML)
Data Storage	Stores data in tables (rows and columns).	Uses dataset (rows and columns) for training models.
Data Cleaning	Removes duplicates and errors (clean data).	Needs clean data, otherwise results are not accurate.
Data Retrieval	Uses SQL queries to retrieve data fast and easy.	Uses simple algorithms to access and use datasets.
Integration	Data is error-free and reliable.	Uses integrated data for strong model training.
Scalability	Can handle small to very large data.	Works with thousands to millions of records.
Real-Time Processing	Real-time storage and retrieval support.	Needs real-time data to learn and update models.
Analytical Tools	Provides SQL and other tools for analysis.	Uses database data and tools for faster results.

4.3 Stages of the Data Science Life Cycle

What is Data Science Life Cycle?
The Data Science Life Cycle is a step-by-step process used to solve real-world problems using data. Just like a recipe has steps to make food, data science has stages to get useful results from data. Each stage is important and helps move towards the final goal.

---

1. Problem Definition

- **What is it?**
  - This is the first and most important step. Here, we try to understand exactly what problem we are trying to solve. We talk to the people who

have the problem (like a business or a teacher) and ask questions to make sure we know what they want.

- **Setting Goals:**
  - We decide what success will look like. These are called **Key Performance Indicators (KPIs)**. KPIs are clear, measurable targets. For example, "increase sales by 15% in two months" or "reduce website loading time by 20% in 30 days."
- **Scope and Limitations:**
  - We also decide what is included in the problem and what is not. This helps us stay focused and not waste time on things that don't matter.
- **Outcome:**
  - At the end of this stage, we have a clear question or hypothesis to work on. This helps guide all the next steps.

---

2. Data Collection

- **What is it?**
  - In this step, we gather all the data we need to answer our question or solve our problem.
- **Sources of Data:**
  - Data can come from many places: surveys, experiments, sensors, online databases, or APIs (which are like messengers between different software).
- **Methods:**
  - For simple problems, we might just observe and write down what we see. For more complex problems, we might use online forms, interviews, or special software to collect data.
- **Quality Matters:**
  - The data must be relevant (related to the problem) and accurate (correct). If the data is bad, our results will also be bad.

---

3. Data Cleaning (Data Preprocessing)

- **What is it?**
  - Data cleaning means fixing the data so it is ready for analysis. This is also called data preprocessing.
- **Steps Involved:**
  - Remove errors (like wrong or impossible values).
  - Delete duplicate entries (same data written more than once).
  - Fill in missing values with best guesses or averages.
  - Sometimes, create new columns from existing data (for example, making a "day of the week" column from a date).
- **Why is it important?**
  - Clean data is very important because if the data is messy, our analysis and results will be wrong. There is a saying: "Garbage in, garbage out."

4. Data Analysis

- **What is it?**
  - Now, we study the cleaned data to find patterns, trends, or important facts.
- **Techniques Used:**
  - We use statistics (like averages, percentages) or machine learning techniques to help us understand the data.
- **Visualization:**
  - We use graphs, charts, and tables to make the data easier to understand and to spot patterns quickly.
- **Testing Ideas:**
  - This step helps us test our early ideas and see if there are any interesting relationships in the data. For example, does weather affect sales?

---

5. Data Modeling

- **What is it?**
  - In this stage, we organize the data into a model or structure that helps us make predictions or understand relationships.
- **Creating Models:**
  - We create diagrams (like ER diagrams) to show how different things (like customers and products) are related.
  - We decide how to store and get the data efficiently, often using databases.
- **Testing the Model:**
  - We test the model with old (historical) data to see if it works well for predictions or analysis.
- **Goal:**
  - The goal is to have a system that is organized, efficient, and ready for making predictions.

---

6. Model Evaluation

- **What is it?**
  - Here, we check how good our model is at solving the problem.
- **Measuring Performance:**
  - We measure its accuracy (how often it is correct) and reliability (does it work every time?).
  - We check if the model meets the KPIs we set earlier.
- **Improvement:**
  - If the model is not good enough, we go back and improve it. Sometimes, we need to collect more data or clean it better.
- **Safety and Privacy:**

- We also make sure the model is safe and respects privacy, especially if it uses sensitive data.

---

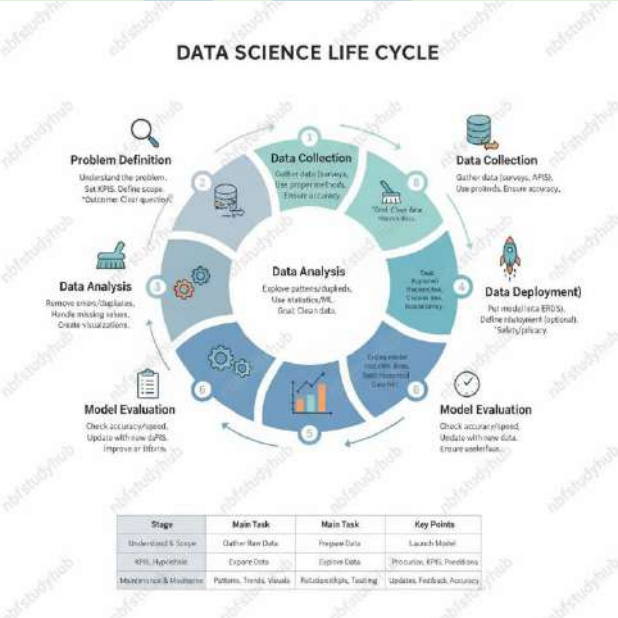
7. Model Deployment

- **What is it?**
  - When the model is ready and works well, we put it into use. This is called deployment.
- **How is it done?**
  - The model is added to a website, app, or database so it can work with real data and help real users.
  - Sometimes, we deploy the model in small steps (partial deployment) to test it before a full launch.
- **After Deployment:**
  - The model starts making predictions or giving insights in the real world.

---

8. Maintenance and Monitoring

- **What is it?**
  - After deployment, we keep checking the model to make sure it works well over time.
- **What do we check?**
  - We monitor its accuracy, speed, and reliability.
  - If there are problems or if the data changes, we update the model.
  - User feedback is also used to make improvements.
- **Why is it important?**
  - This step is ongoing, so the model stays useful and up-to-date. If we ignore this step, the model can become less accurate as time passes or as new data comes in.



Summary Table

Stage	Main Task	Key Points
Problem Definition	Define the problem and set goals (KPIs)	Clear question, measurable targets
Data Collection	Gather data from various sources	Data must be relevant and accurate
Data Cleaning	Fix and organize data	Remove errors, fill missing values
Data Analysis	Study data for patterns and trends	Use statistics, graphs, and charts
Data Modeling	Build a structure/model for the data	Organize data, test with old data
Model Evaluation	Check model's performance	Measure accuracy, improve if needed
Model Deployment	Put the model into real use	Add to website/app, start real predictions
Maintenance & Monitoring	Keep checking and improving the model	Update as needed, use feedback

THE END