Decision Trees

Overview

A **Decision Tree** is a machine learning algorithm that makes predictions by asking a series of yes/no questions.
It resembles a flowchart, where each internal node represents a decision based on a feature, and each leaf node represents an outcome.

Decision Trees are easy to understand and interpret — one of the reasons they are often used in business and healthcare systems.

Key Concepts

- **Root Node:** The starting point where the first decision is made (e.g. "Is income > £50,000?").
- **Branches:** Represent possible outcomes or paths (e.g. *Yes* → *Approve*, *No* → *Decline*).
- **Leaf Nodes:** The final output or prediction.
- **Splitting:** The process of dividing data into subsets based on certain conditions.
- **Information Gain / Gini Impurity:** Metrics that measure how well a split separates the data.

Visual Summary

Start ├── Income > £50,000 → Approve Loan └── Income ≤ £50,000 → Deny Loan Each "branch" leads to another question or to a final decision.

Real-World Applications

- 🏦 **Banking:** Approving or rejecting loan applications.
- 🩺 **Healthcare:** Diagnosing diseases based on symptoms.
- 🛍️ **E-commerce:** Recommending product categories to customers.
- 💁 **Customer Support:** Creating intelligent troubleshooting flows.

Example

A bank might use a Decision Tree to determine credit eligibility:

IF income > £50,000 AND credit_score > 700 → APPROVE ELSE → REJECT

Why It Matters

Decision Trees are transparent — users can easily follow the logic behind a prediction.
They are often used as the building blocks for **Random Forests** and **Gradient Boosted Trees**, which combine many trees to achieve greater accuracy.