



ROMENTER AI

# MODULE 01 // EXECUTIVE BRIEFING

## Supervised Learning Guide

Prepared By :

**Ben Jerome**

Founder of Romenter.ai

[www.romenter.ai](http://www.romenter.ai)

When we talk about "Machine Learning, " we are usually discussing one of two distinct approaches.



The difference lies entirely in whether we give the computer the "correct answer" during training.

Supervised Learning: We teach the AI using historical examples where the answer is known.

Unsupervised Learning: We give the AI raw data with no answers and ask it to find structure on its own.

Below is an explanation of these concepts using the workflows of your commercial team.

### **1. Supervised Learning:**

The Graduate QS: The Analogy: Training a Junior Quantity Surveyor using past projects.

Imagine you hire a fresh graduate. You sit them down with the "As-Built" records of your last 50 projects. You hand them the drawings for Project #1 and say, "Estimate the concrete volume. " The Graduate guesses: "1,000 m<sup>3</sup> . "You check the Final Account (the Label): "Wrong. The actual was 1,200 m<sup>3</sup> . "The Correction: The graduate adjusts their mental model. "Okay, I missed the waste factor and the deep foundations. "They repeat this 50 times. By the end, they have learned the rules that map drawings to costs.

How the AI Works (Labeled Data):In Supervised Learning, the data is "Labeled. " We have an input (the drawing) and a confirmed output (the final cost).

The AI tries to predict the output, compares its guess to the actual answer, and adjusts its internal math to minimize the error.

Best Use Cases in Construction: Cost Prediction: "Here are 100 past hospital projects (Inputs) and their final costs (Labels).

Predict the cost of this new hospital. "RFI Classification: "Here are 5,000 past RFIs. We have labeled them as either 'MEP' , 'Structural' , or 'Architectural'.

Learn to route this new RFI to the right team. "Safety Monitoring: "Here are 1,000 photos of workers. We have drawn a box around every person not wearing a hard hat.

Learn to spot the missing hat. "Key Limit: It can only predict what it has seen before. If you trained the Graduate only on hospitals, they will fail at estimating a bridge.

DATA SOVEREIGNTY: Public LLMs (like ChatGPT) learn from your uploads. Romenter isolates your data. When you upload drawings to the Insight Engine, they are vectorized locally within your secure partition. Your intellectual property never trains the public model.

2. Unsupervised Learning: The Forensic Auditor: The Analogy: Letting a Forensic Auditor loose in a messy file room.

Imagine you suspect collusion in your supply chain, but you don't know who is involved or how they are doing it. You hire an external forensic auditor and dump 10,000 raw invoices on their desk. You provide no answer key. You don't tell them what "fraud" looks like.

You just say, "Group these documents and tell me if anything looks weird. "The Auditor starts sorting. "I've noticed that Vendor A and Vendor B always submit invoices on the same day. "I've noticed that Vendor C's invoices are formatted exactly like Vendor D's, down to the pixel.

"I've found a cluster of invoices that are all just under £50,000 (the approval limit).

"How the AI Works (Unlabeled Data): In Unsupervised Learning, the data is "Unlabeled. " The AI doesn't know what "fraud" or "collusion" is. It simply looks at the mathematical properties of the data and groups similar items together (Clustering) or identifies items that don't fit any group (Anomaly Detection).

Best Use Cases in Construction: Fraud Detection: "Here are all our transactions. Find the outliers that don't fit the standard pattern. "Vendor Segmentation: "Group our 5,000 suppliers based on their performance, location, and pricing behavior.

Tell us which groups exist.

"Schedule Optimization: "Look at these 100 schedules. Find common clusters of activities that usually happen together, even if we didn't explicitly link them. "Key Benefit: It can find patterns you didn't know existed.

The human didn't know to look for "invoices under £50k, " but the AI found the cluster anyway.

Comparison: When to Use Which?

Feature: Supervised Learning (The Graduate)

Unsupervised Learning (The Auditor)

Data Required Labeled Data (Input + Correct Answer). Hard to get. Requires human effort to tag history.

Raw Data (Input only).

Easy to get. Just dump the database.

The Goal Prediction. "What will the cost be?

"Discovery. "What is happening in my data?"

The Analogy Following a textbook / Mentorship. Detective work / Sorting a messy room.

DATA SOVEREIGNTY: Public LLMs (like ChatGPT) learn from your uploads. Romenter isolates your data. When you upload drawings to the Insight Engine, they are vectorized locally within your secure partition. Your intellectual property never trains the public model.



**Risk:** If the historical data is biased (e.g., all past projects were over budget), the AI will predict over-budget as "normal. "It might find patterns that are useless (e.g., "All these invoices are printed on white paper").

**Conclusion:** Use Supervised Learning when you want to automate a task you already know how to do (e.g., classifying emails, estimating costs), provided you have clean historical records to train the model.

Use Unsupervised Learning when you want to explore the unknown (e.g., finding new types of risk, segmenting bad actors) where you don't have a pre-defined list of "answers. "

DATA SOVEREIGNTY: Public LLMs (like ChatGPT) learn from your uploads. Romenter isolates your data. When you upload drawings to the Insight Engine, they are vectorized locally within your secure partition. Your intellectual property never trains the public model.



ROMENTER AI

THEORY INTO PRACTICE  
LAUNCH INSIGHT ENGINE