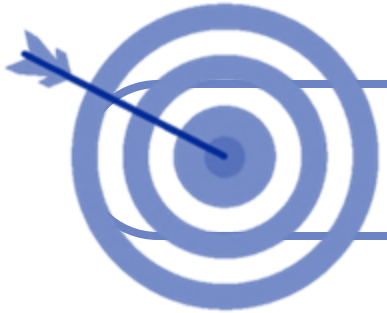
ADNOC Accelerator Programme

Artificial Intelligence

COHORT 2

Machine Learning Fundamentals

Machine Learning Fundamentals

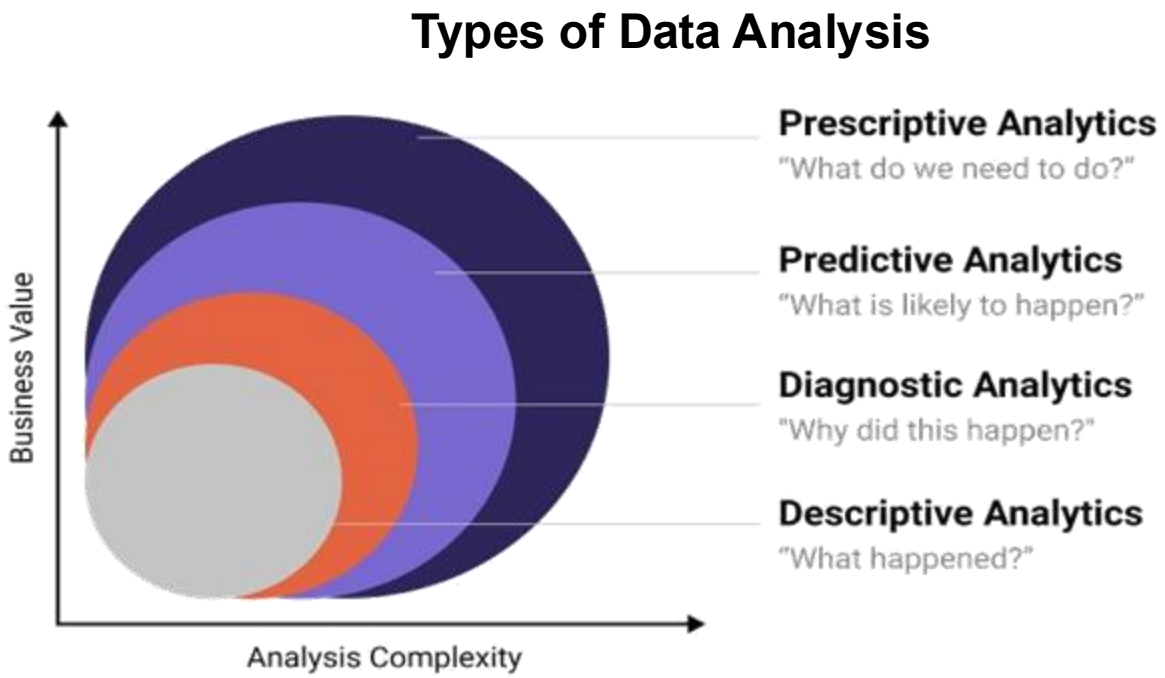


LEARNING OBJECTIVES

1. Understand machine learning basics
2. Explore supervised and unsupervised learning applications
3. Determine which ML model to use

Machine Learning allows systems to continuously improve

Machine Learning (ML) is a subfield of Artificial Intelligence where system learn from data and improve over time without being explicitly programmed



ML mostly helps business with "Predictive" & "Prescriptive" analytics

ML models can identify patterns to make predictions or decisions

The models analyse large datasets to find patterns in two key ways

Supervised Learning

Learning by example

1. **Regression**
2. **Classification**

Unsupervised Learning

Learning by observation

3. **Clustering**

ML models can identify patterns to make predictions or decisions

The models analyse large datasets to find patterns in two key ways

Supervised Learning

Learning by example

1. **Regression**
2. **Classification**

Unsupervised Learning

Learning by observation

3. Clustering

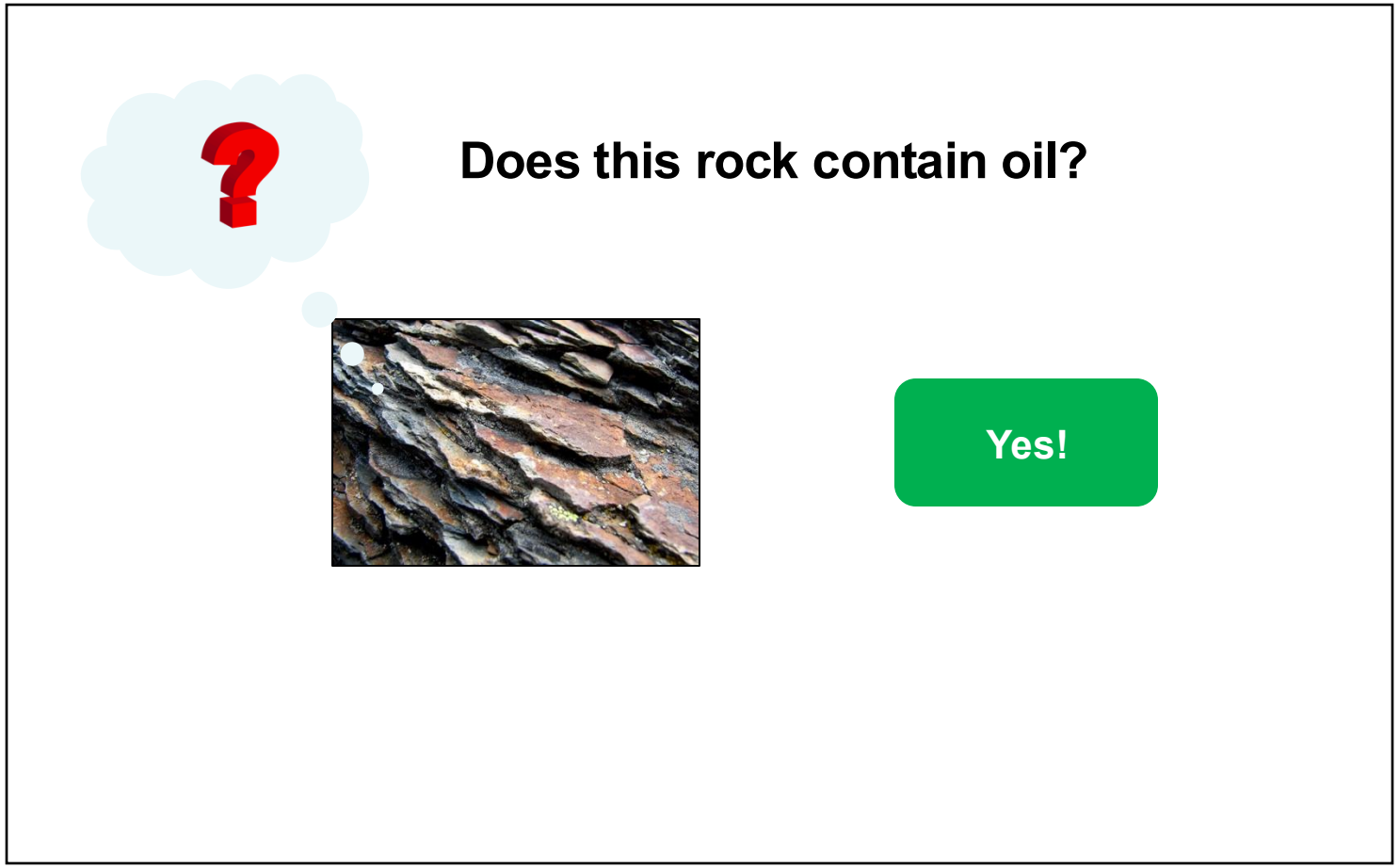
Supervised models learn from labelled data

1 Contains Oil

2 Doesn’t Contain Oil

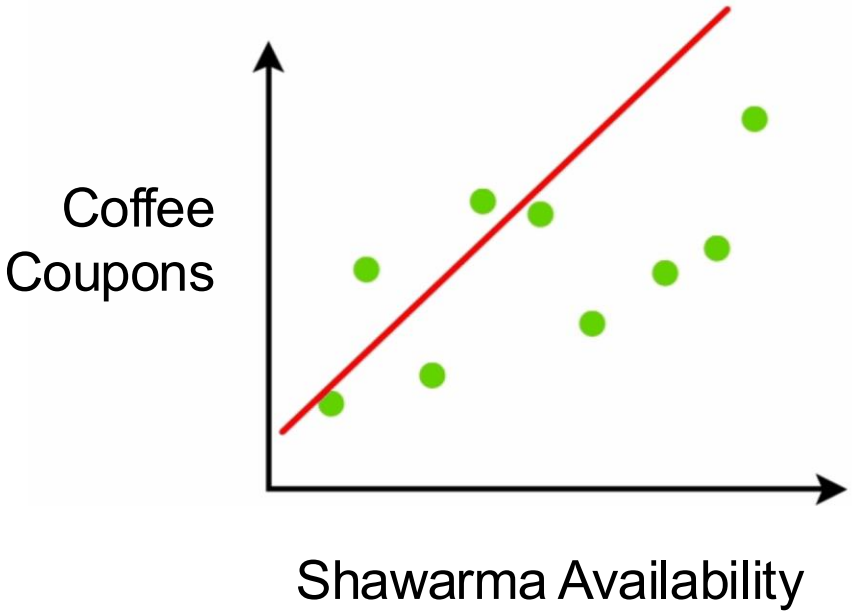
3 Contains Oil

Supervised models learn from labelled data



Linear Regression predicts the relationship between variables

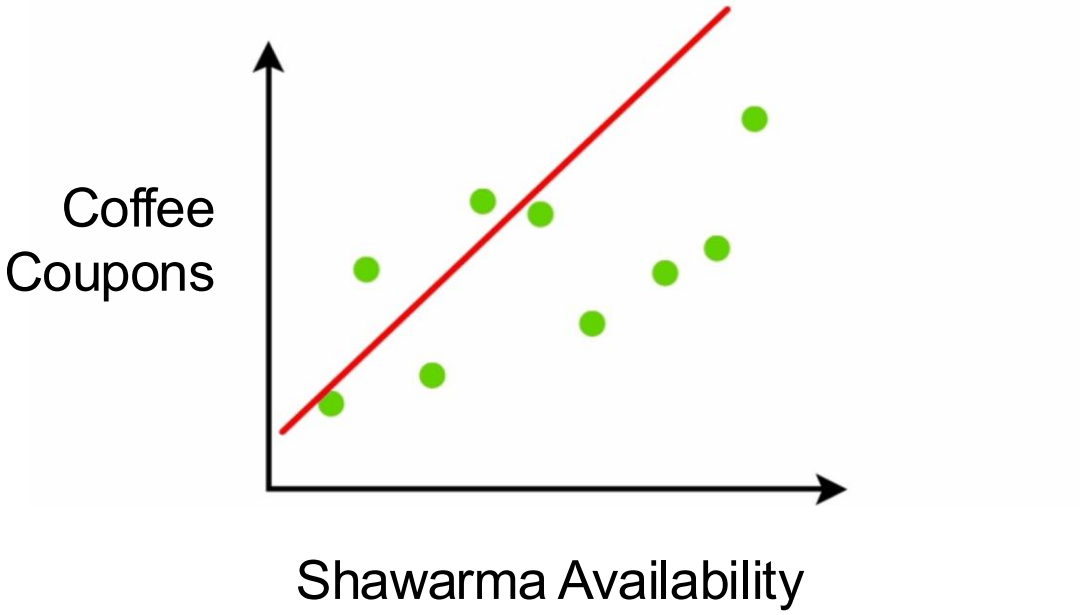
Canteen Satisfaction ~ coffee coupons + shawarma availability



Source: Medium

Linear Regression predicts the relationship between variables

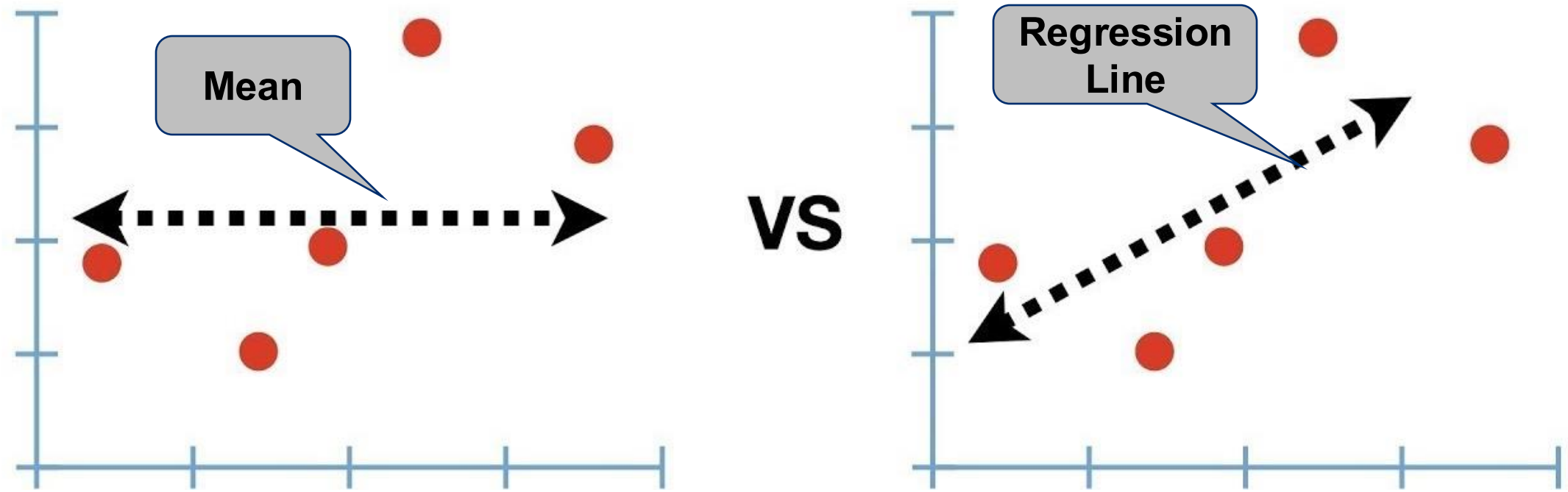
Canteen Satisfaction = 3 + 2 x coffee coupons + 1.5 x shawarma availability



Source: Medium

R^2 shows how much of model is explained by chosen variables

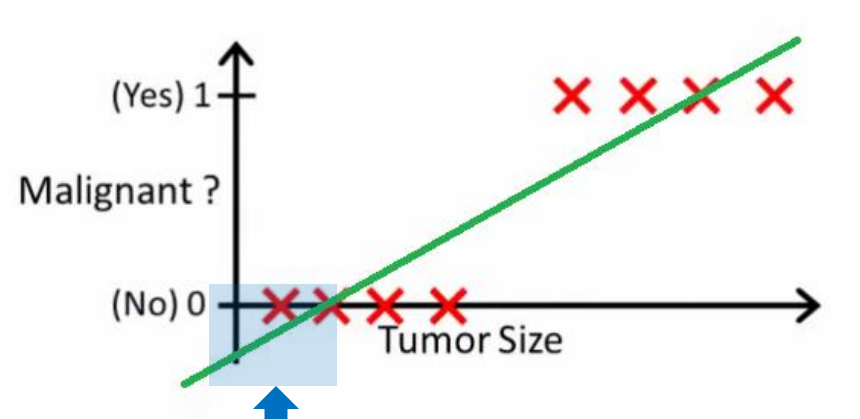
R^2 measures how much better your model is at predicting compared to just using the mean.



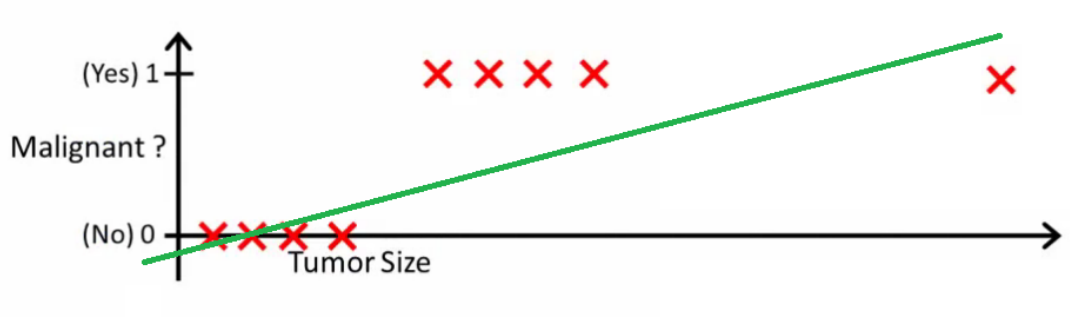
Making It Happen

Train Model → Make Predictions

Linear Regression can cause issues when used for Classification



Negative Values

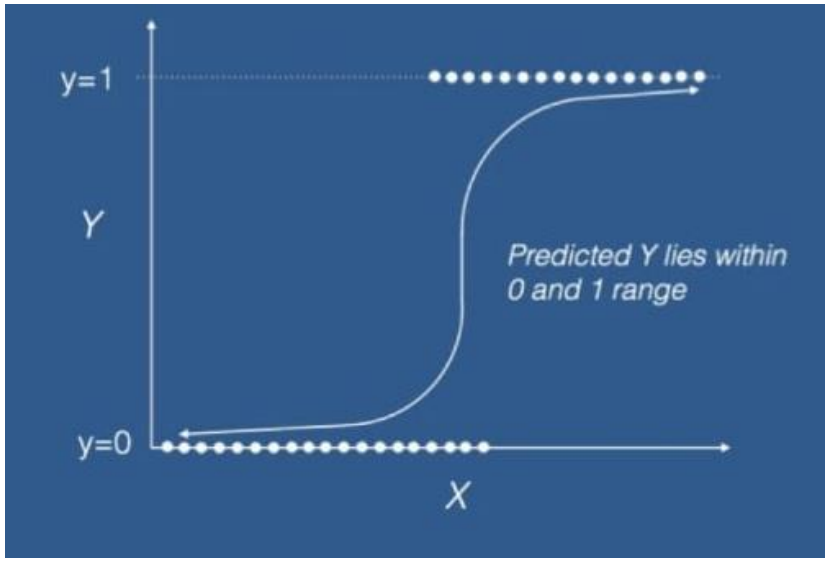


Sensitivity to Outliers

Logistic Regression works when the outcomes are binary

Can you predict whether an equipment will fail within the next two months, caused by factors like wear and tear?

Boundaries change to 0 and 1



This yes/no questioning helps the model ascertain whether something belongs to a particular class or not

Which of these is a binary outcome?

A. Spam versus Not Spam

B. Age of Employee

C. Number of Holidays

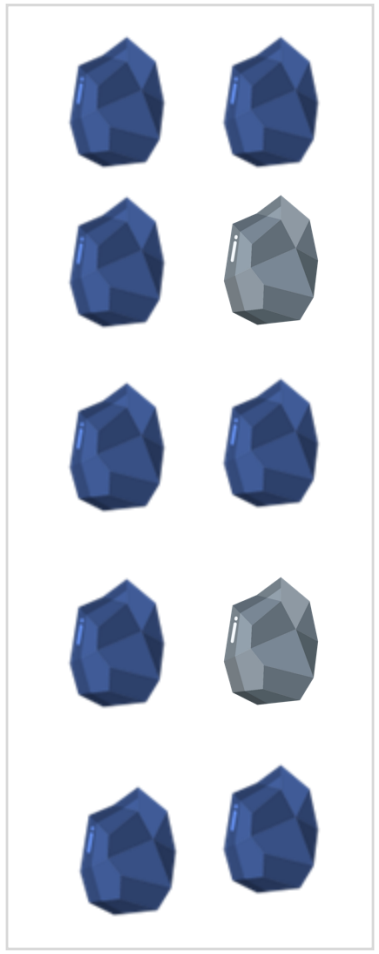
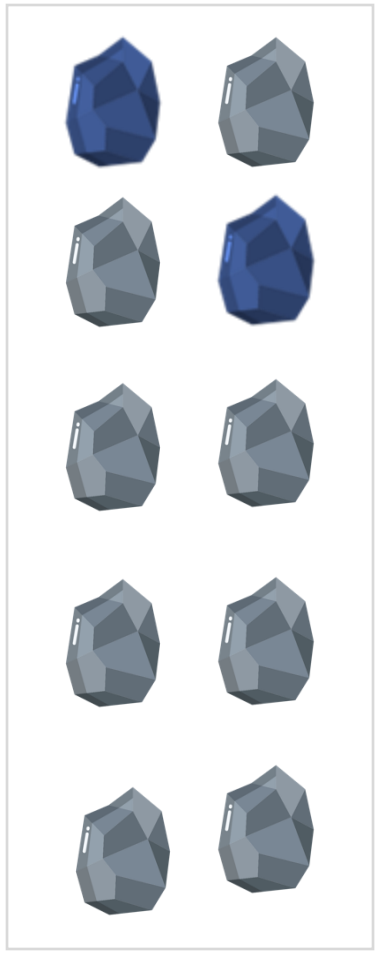
Which of these is a binary outcome?

A. Spam versus Not Spam

B. Age of Employee

C. Number of Holidays

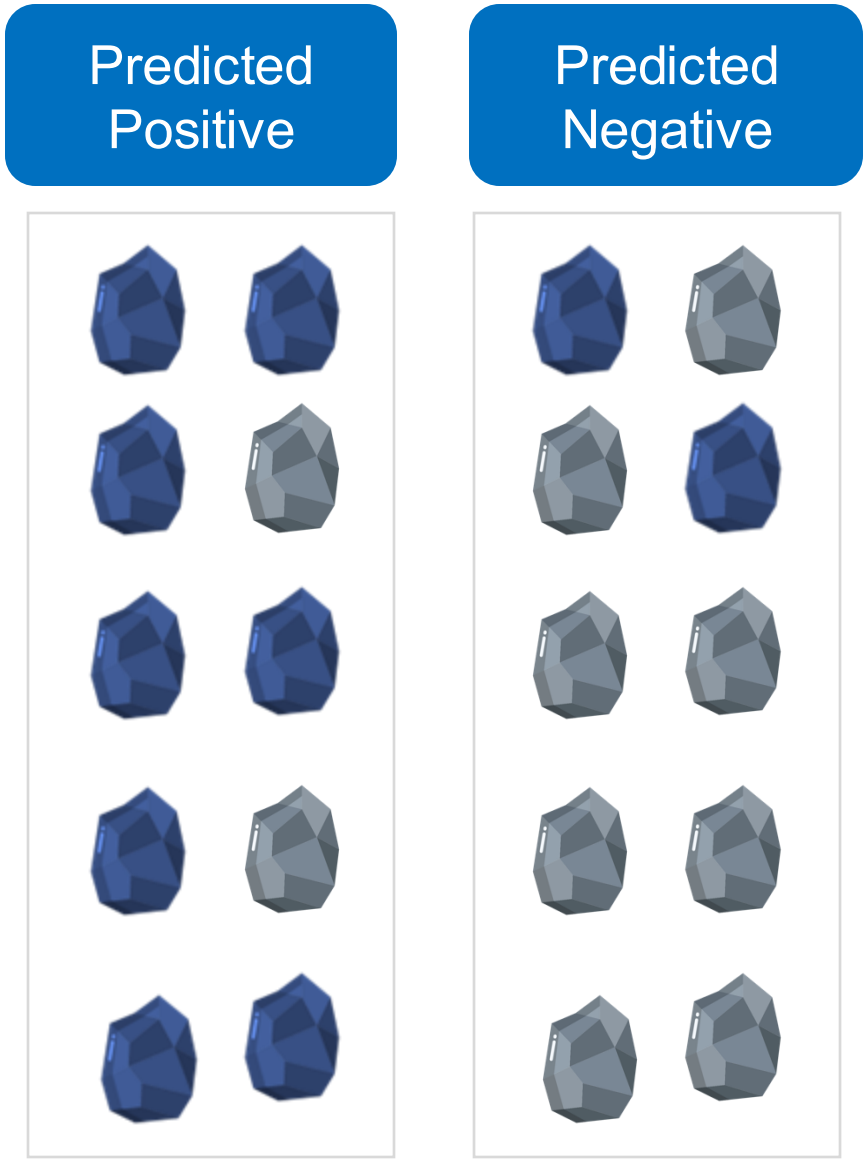
Confusion Matrices help to evaluate the model

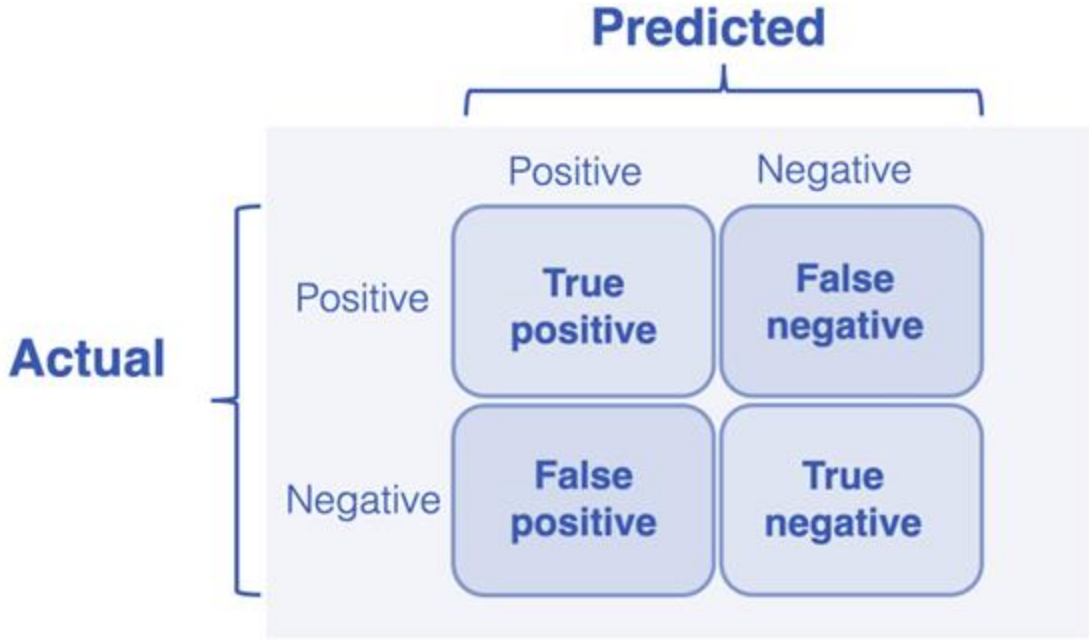
	Predicted Positive	Predicted Negative
Actual		

Blue rocks contain oil

Grey rocks don’t contain oil

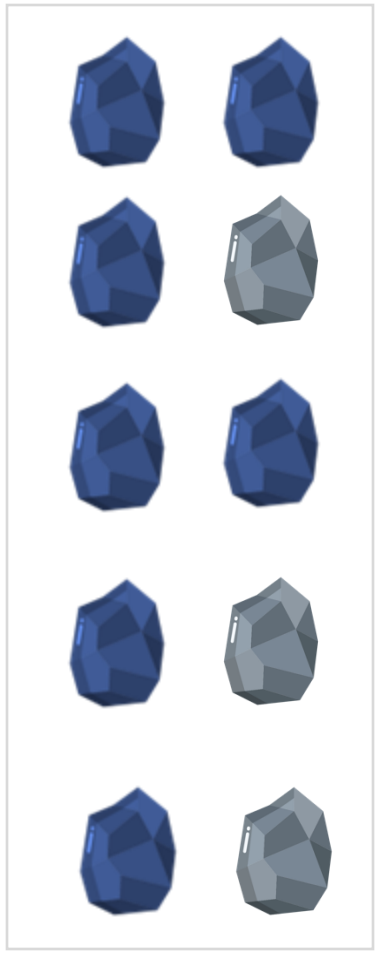
Confusion Matrices help to evaluate the model

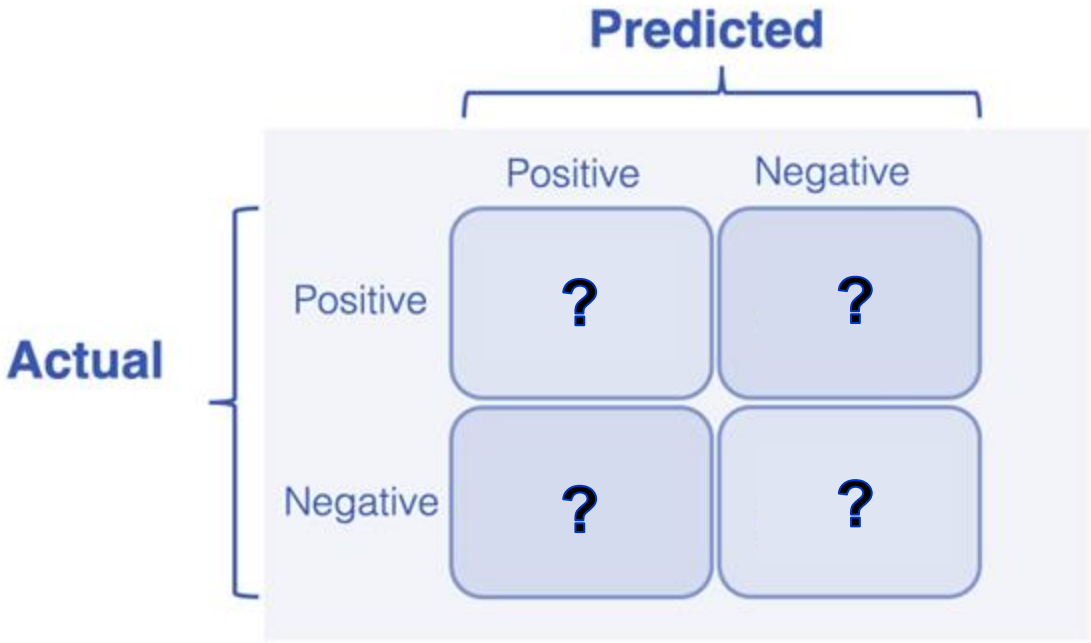




Predicted positives can be true or false in actual

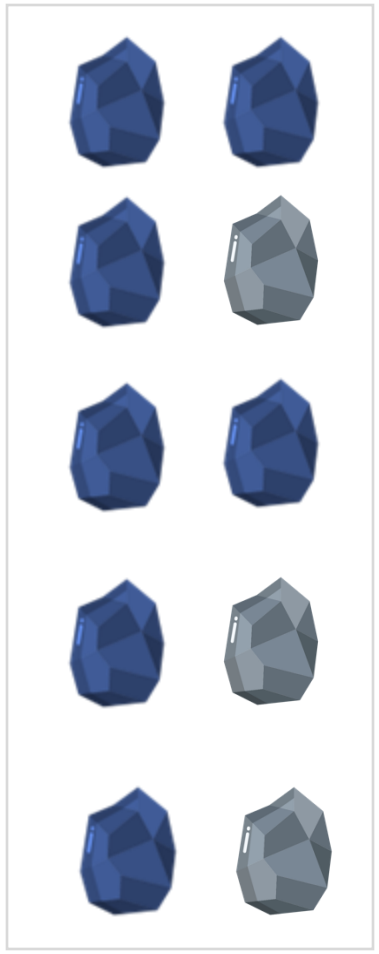
Predicted Positive

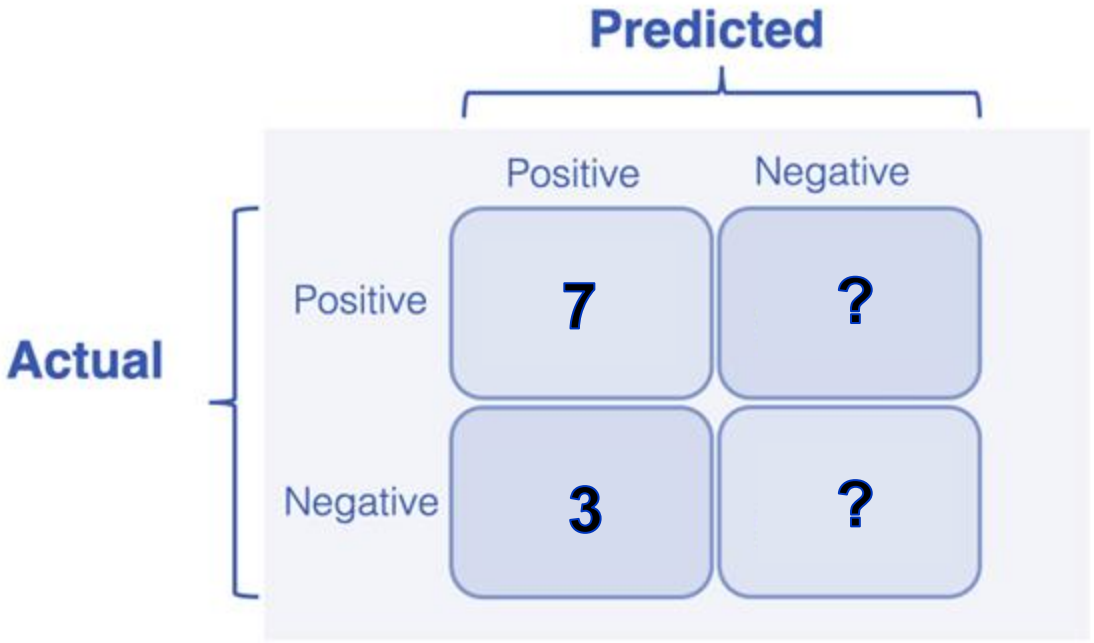




Predicted positives can be true or false in actual

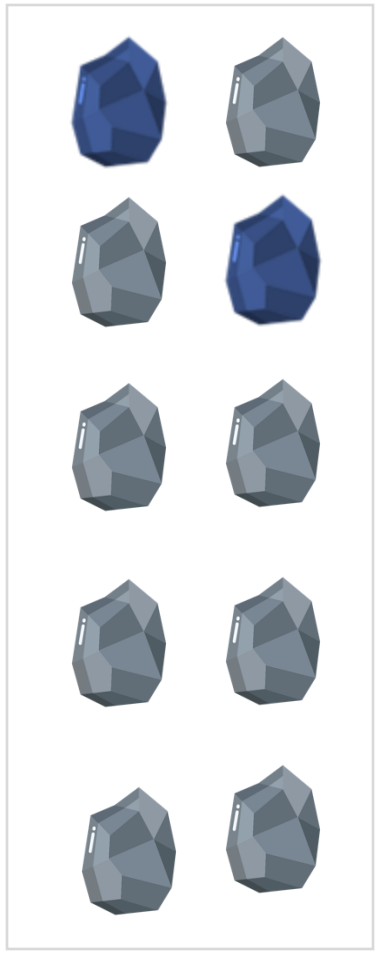
Predicted Positive

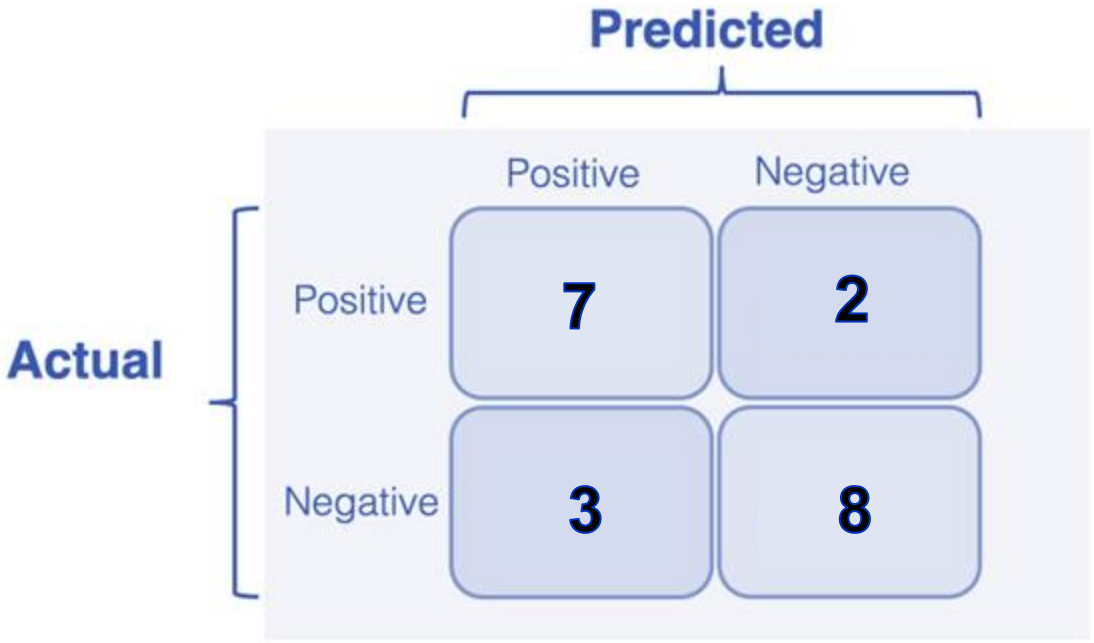




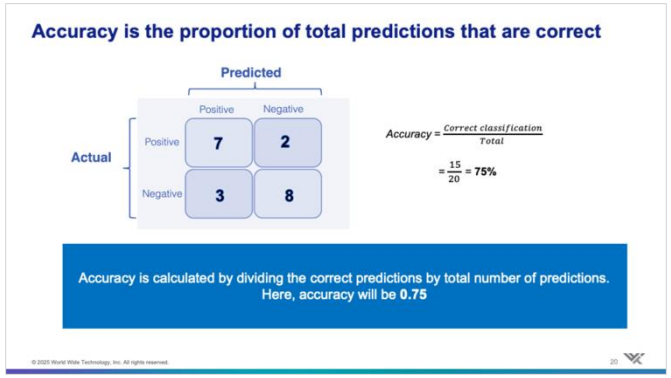
Predicted negatives can be true or false in actual

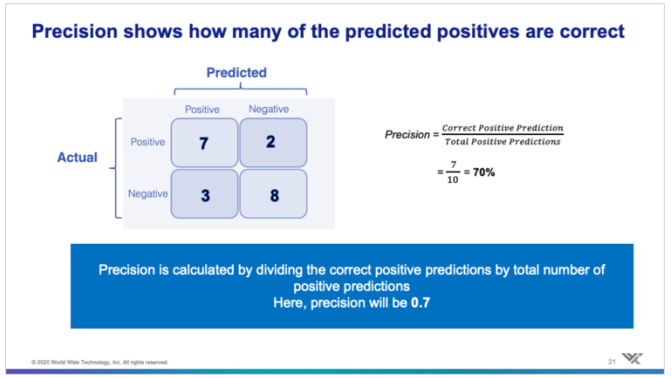
Predicted Negative

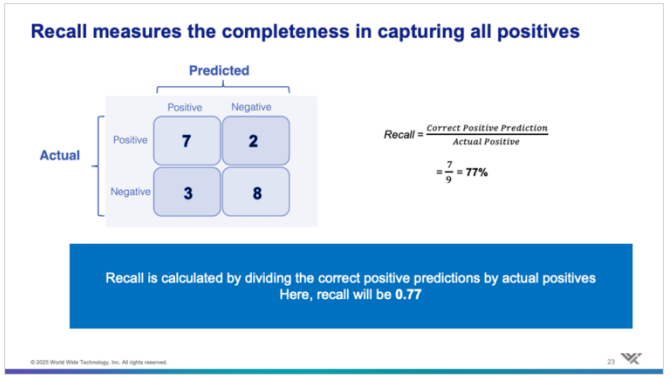




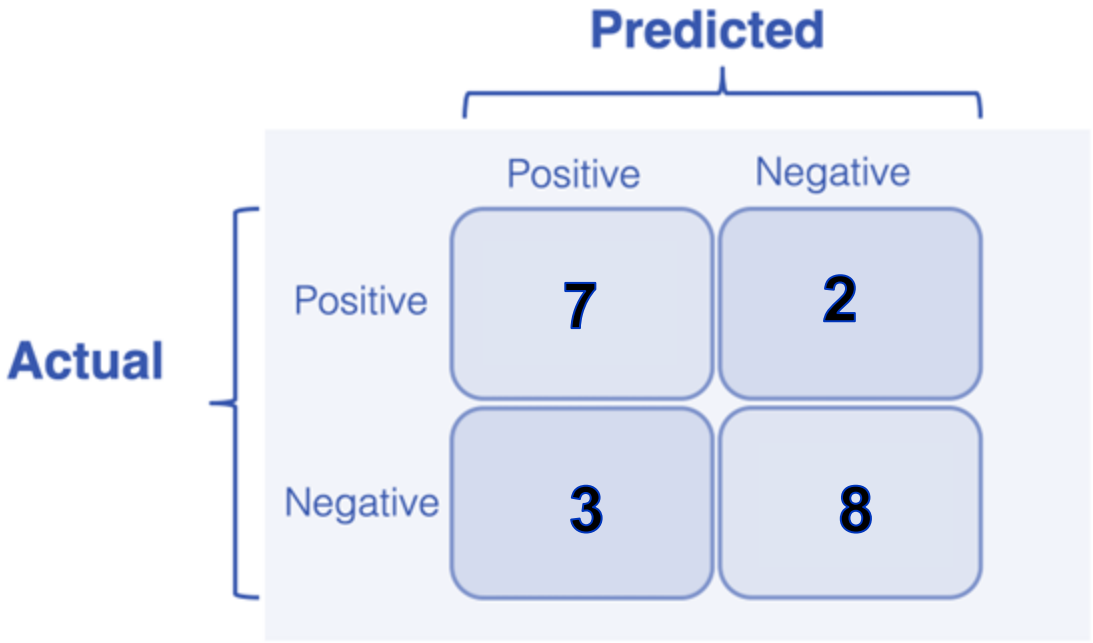
Different metrics can be used to evaluate the model







Accuracy is the proportion of total predictions that are correct

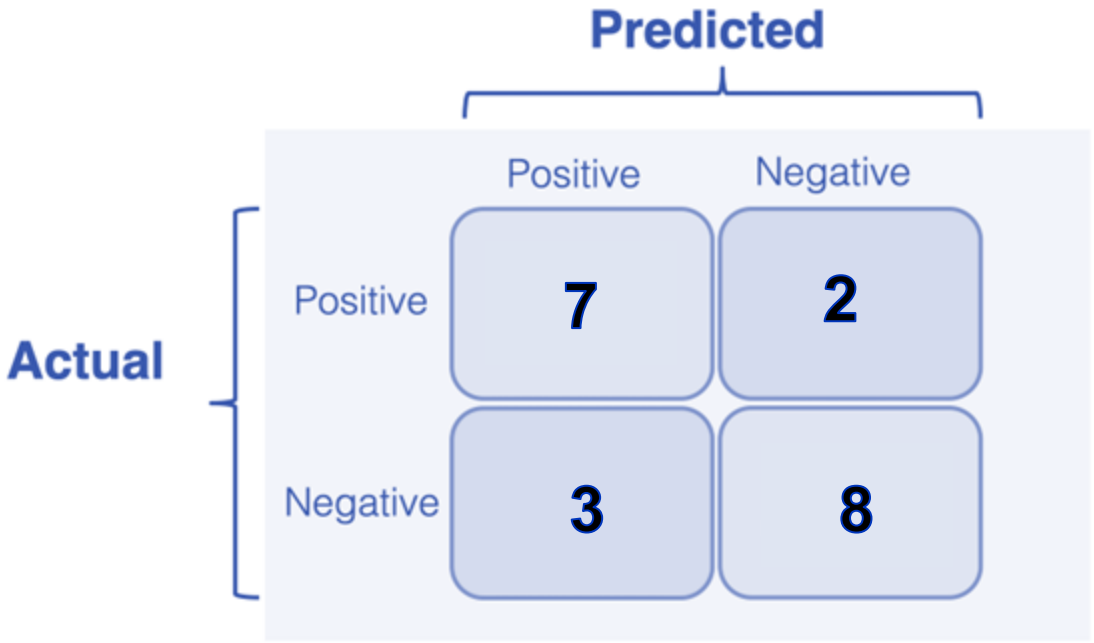


$$Accuracy = \frac{Correct\ classification}{Total}$$

$$= \frac{15}{20} = \mathbf{75\%}$$

Accuracy is calculated by dividing the correct predictions by total number of predictions. Here, accuracy will be **0.75**

Precision shows how many of the predicted positives are correct

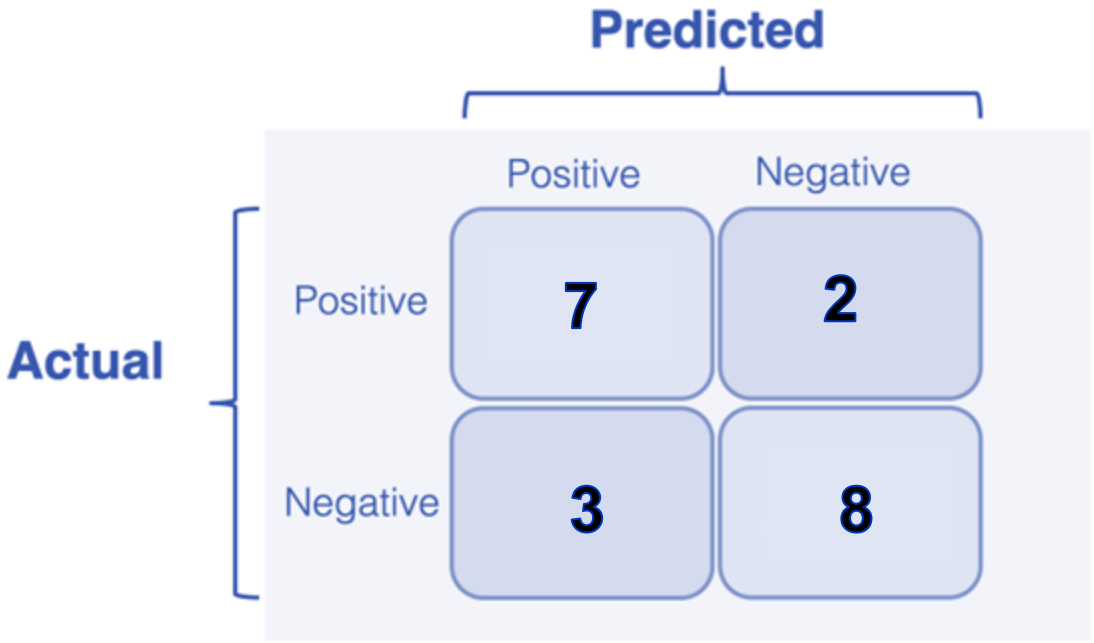


$$Precision = \frac{Correct\ Positive\ Prediction}{Total\ Positive\ Predictions}$$

$$= \frac{7}{10} = \mathbf{70\%}$$

Precision is calculated by dividing the correct positive predictions by total number of positive predictions
Here, precision will be **0.7**

Precision shows how many of the predicted positives are correct



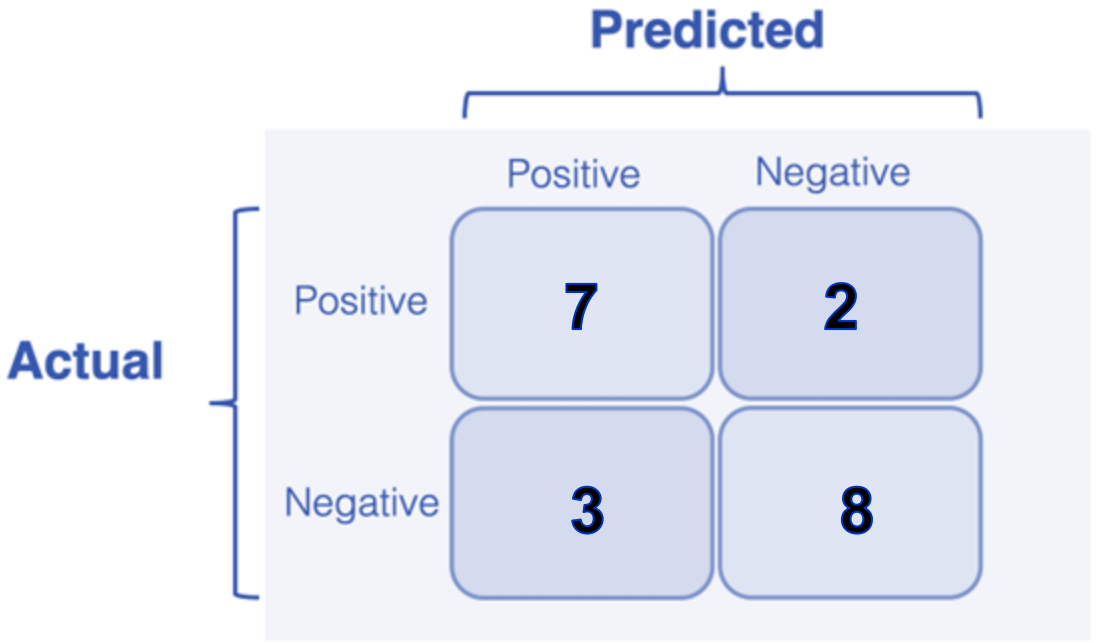
$$Precision = \frac{Correct\ Positive\ Prediction}{Total\ Positive\ Predictions}$$

$$= \frac{7}{10} = \mathbf{70\%}$$

EXAMPLE

Let's say you were a bank handing out loans. You want to ensure you give loans to only those who won't default, so you get that proportion right. Even if you miss some people who won't default, that's not going to be a problem.

Recall measures the completeness in capturing all positives

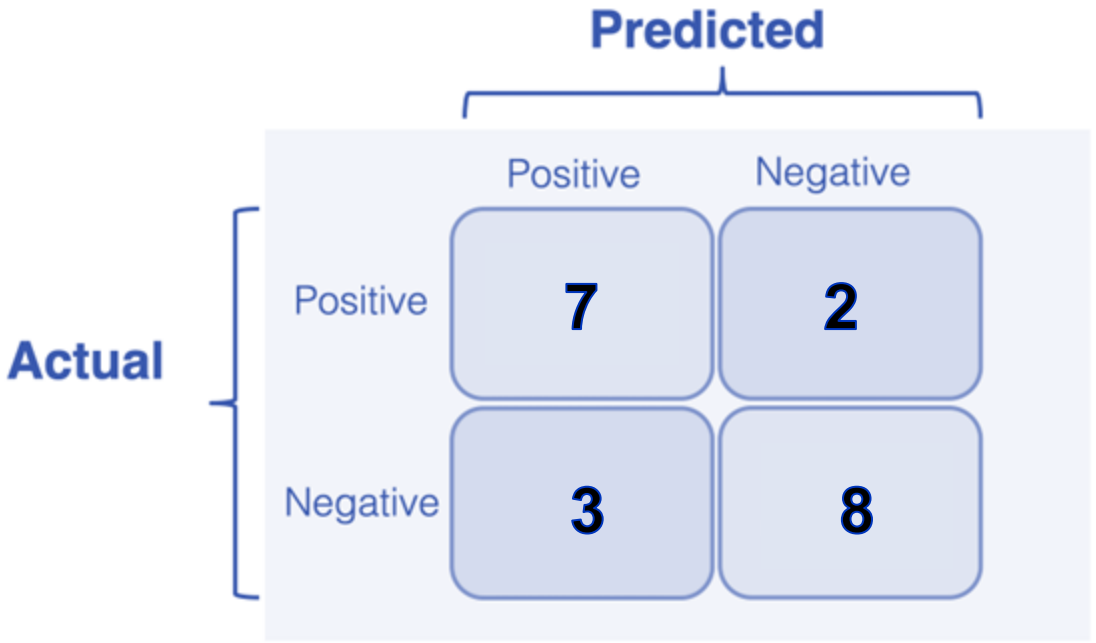


$$Recall = \frac{Correct\ Positive\ Prediction}{Actual\ Positive}$$

$$= \frac{7}{9} = \mathbf{77\%}$$

Recall is calculated by dividing the correct positive predictions by actual positives
Here, recall will be **0.77**

Recall measures the completeness in capturing all positives



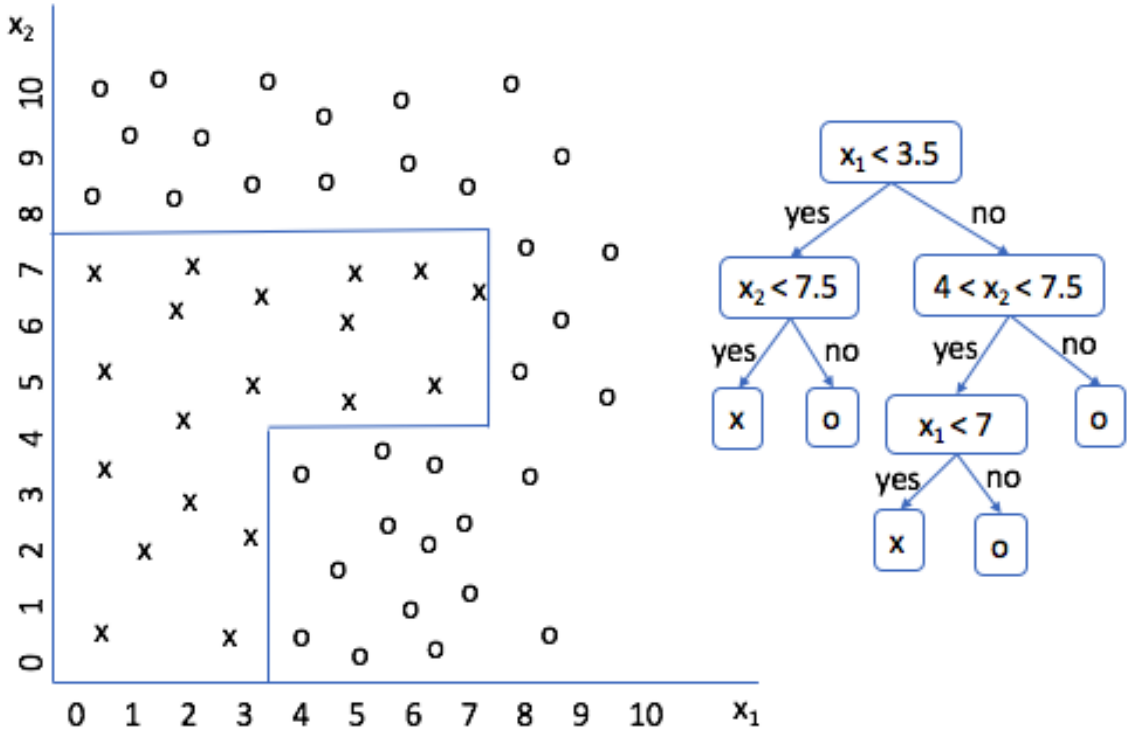
$$Recall = \frac{Correct\ Positive\ Prediction}{Actual\ Positive}$$

$$= \frac{7}{9} = \mathbf{77\%}$$

EXAMPLE

Let's say that a new infectious strain of virus has broken out and you need to quarantine everyone infected so it doesn't spread! Now you would want to make sure all positives are caught. Any positive being left out would cause problems.

Decision Trees investigate data step by step to reach predictions



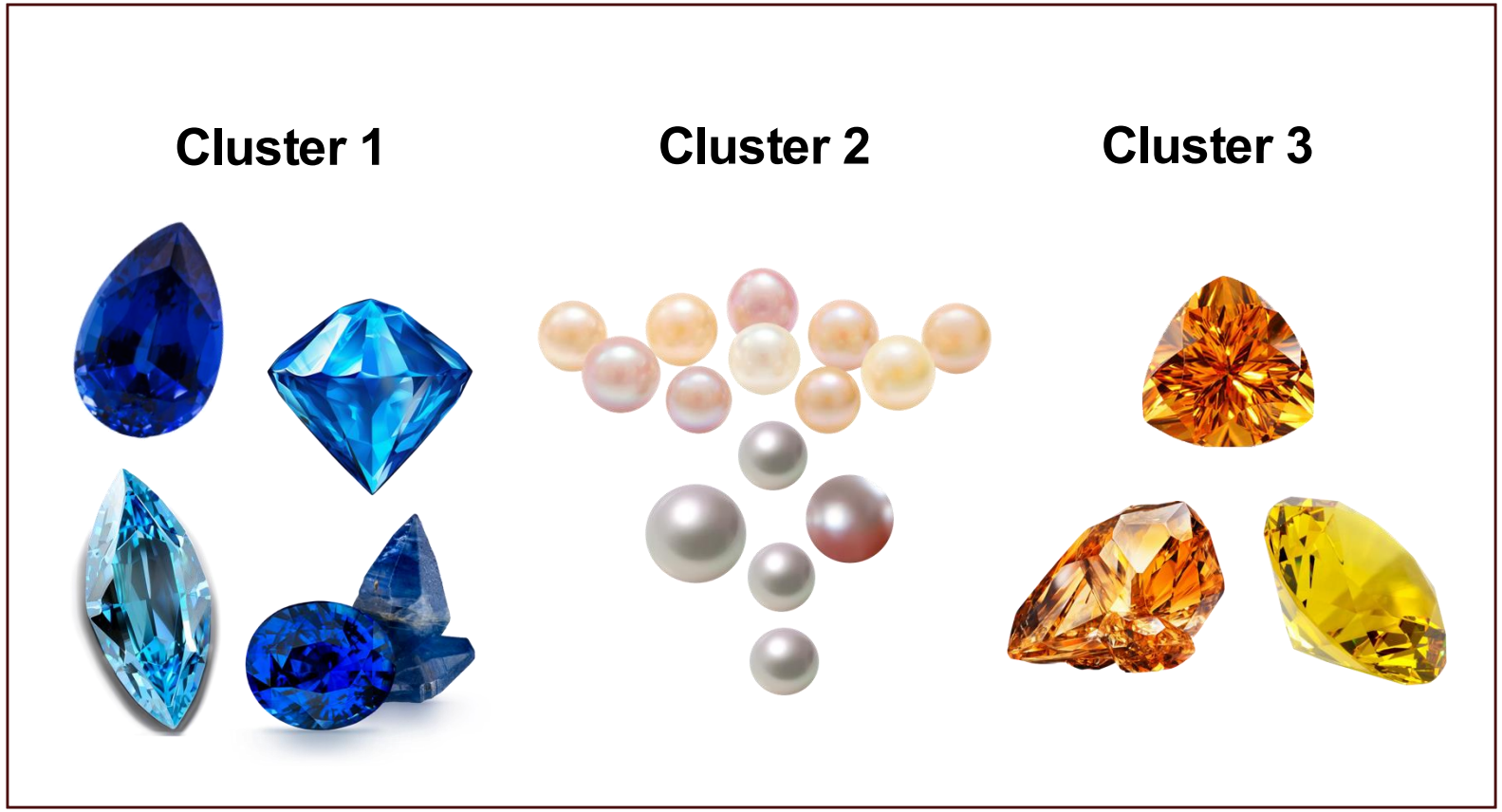
ML models can identify patterns to make predictions or decisions

The models analyse large datasets to find patterns in two key ways

Supervised Learning	Unsupervised Learning
Learning by example	**Learning by observation**
1 Regression	3 **Clustering**
2 Classification	

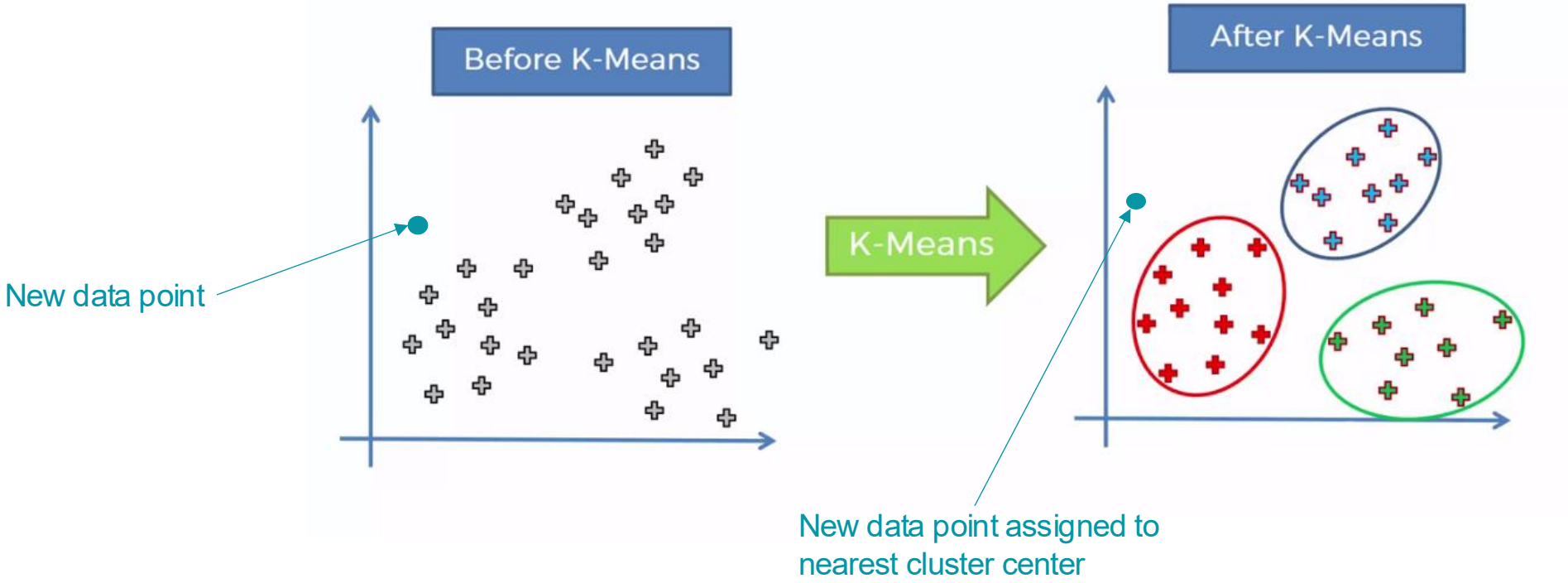
Unsupervised models recognise patterns in un-labelled data

Unsupervised models recognise patterns in un-labelled data



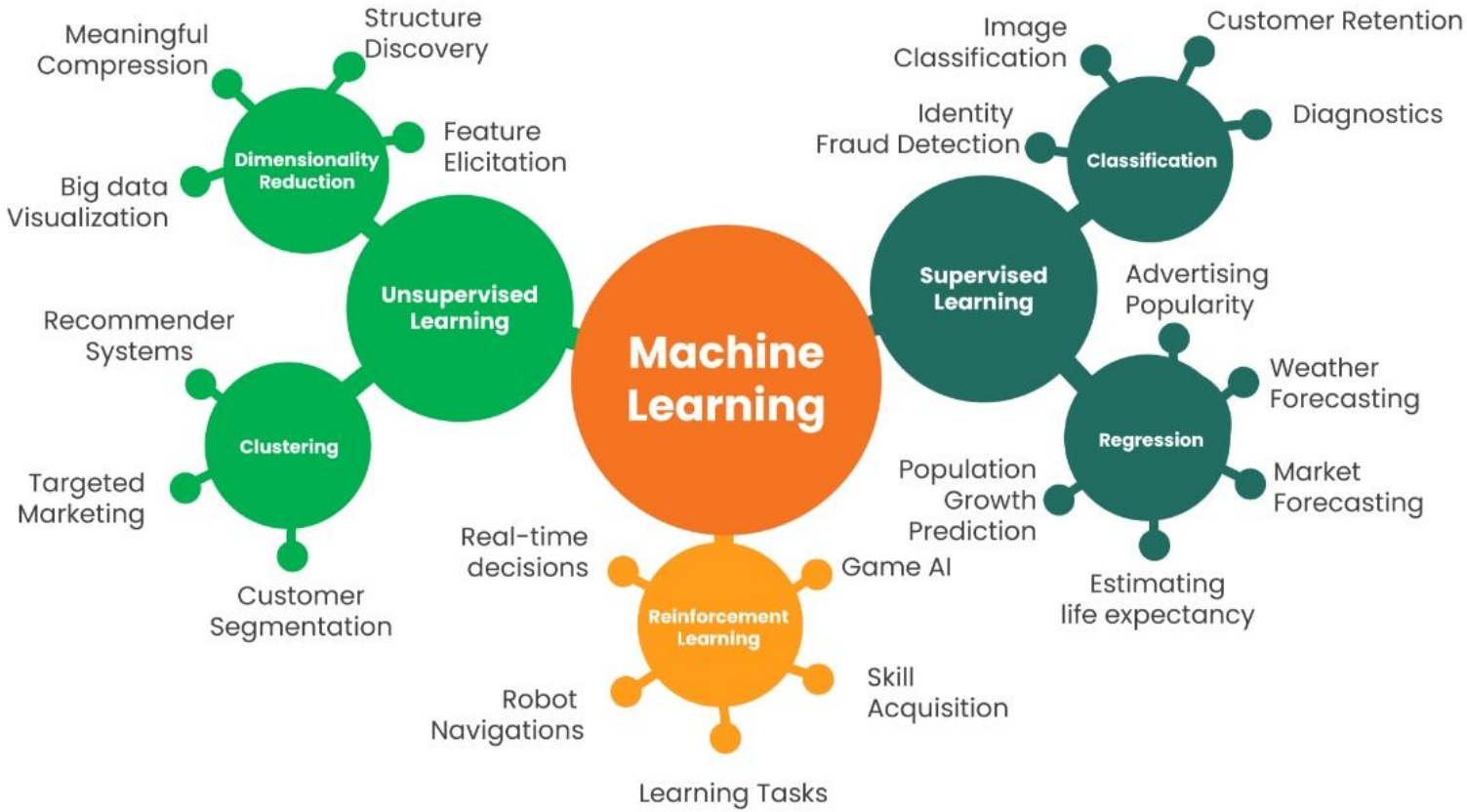
Clustering is used when the data categories are unknown

K-Means: Assign new data to the nearest cluster based on distance to centroids



"Tell me which center you're closest to, and I'll tell you who you belong with. "

The three types of learning can be applied for different use cases



Source: ***Axtria Inc.***

ML model selection should be driven by project-specific needs

No one-size-fits-all approach, but these guidelines help narrow it down

Problem Type
Determines if the model should focus on classification, regression, or clustering

Data Size & Quality
Influences the choice between deep learning and traditional ML

Computational Power
Affects the feasibility of using neural networks vs lightweight models

Interpretability
Guides the choice based on how much understanding the model needs to have

Start by trying a few models, decide which is best for your needs with respect to accuracy, speed and interpretability, then pick!

Machine Learning Fundamentals

In this session, we covered:

- **What machine learning is and how it works**
- **Types and applications of supervised learning**
- **Types and applications of unsupervised learning**
- **What to consider when choosing an ML model**