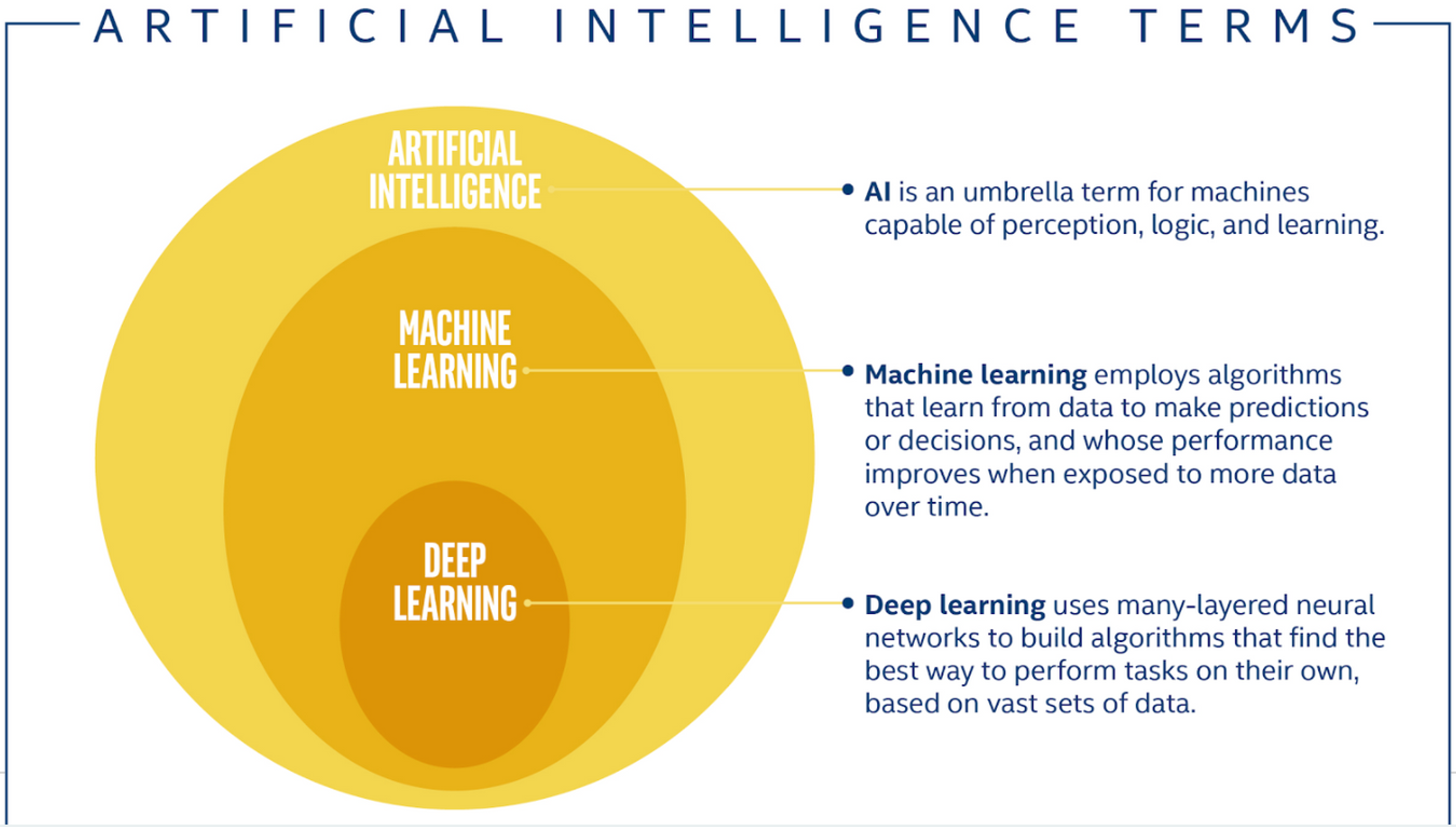
IRON HACK

DATA ANALYSIS BOOTCAMP

MACHINE LEARNING - INTRO

AI / ML / DL

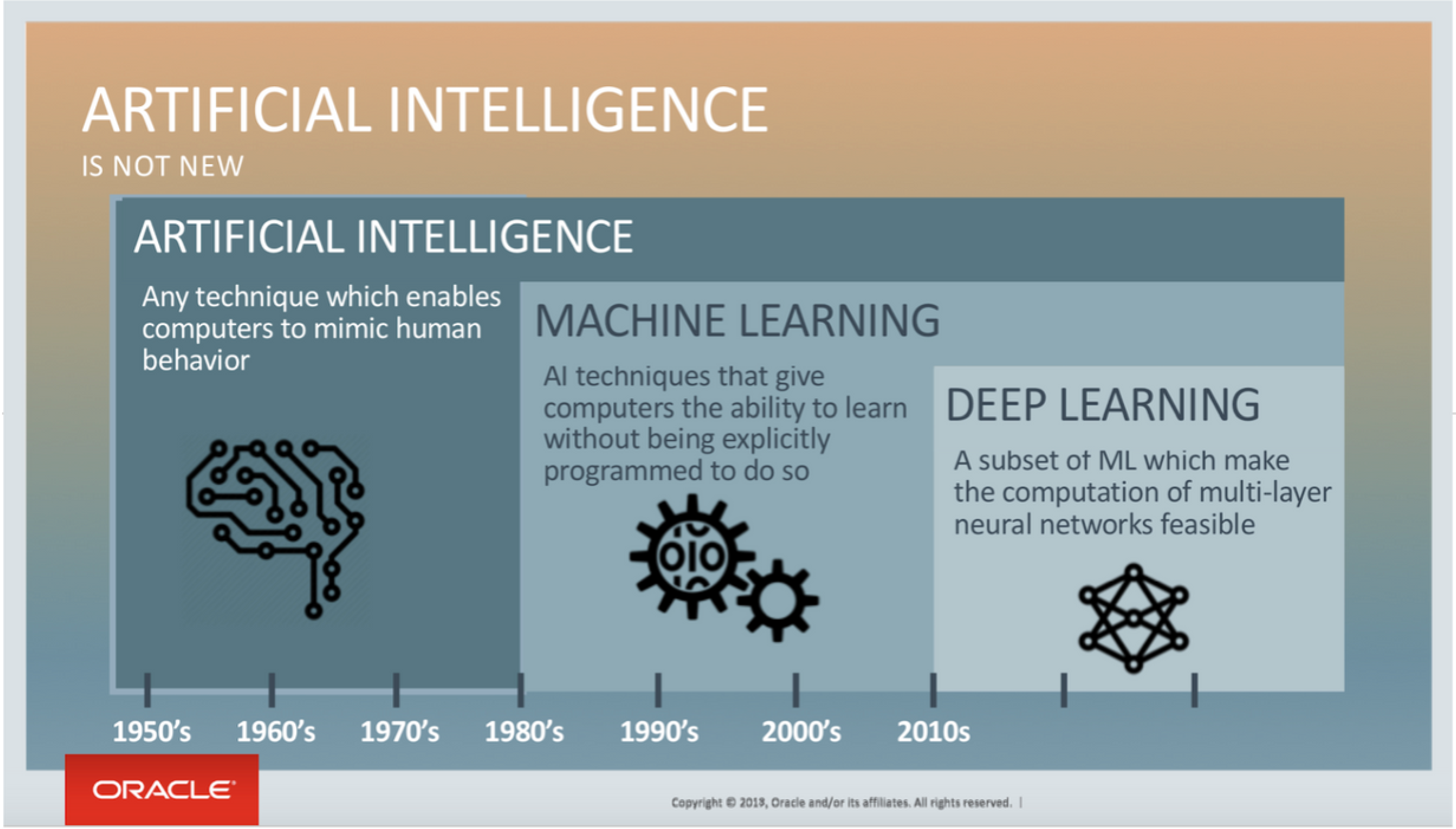


ARTIFICIAL INTELLIGENCE TERMS

- **AI** is an umbrella term for machines capable of perception, logic, and learning.
- **Machine learning** employs algorithms that learn from data to make predictions or decisions, and whose performance improves when exposed to more data over time.
- **Deep learning** uses many-layered neural networks to build algorithms that find the best way to perform tasks on their own, based on vast sets of data.

AI IS NOT NEW

THEN WHY IS IT FAMOUS NOW?



ARTIFICIAL INTELLIGENCE

IS NOT NEW

ARTIFICIAL INTELLIGENCE

Any technique which enables computers to mimic human behavior

MACHINE LEARNING

AI techniques that give computers the ability to learn without being explicitly programmed to do so

DEEP LEARNING

A subset of ML which make the computation of multi-layer neural networks feasible

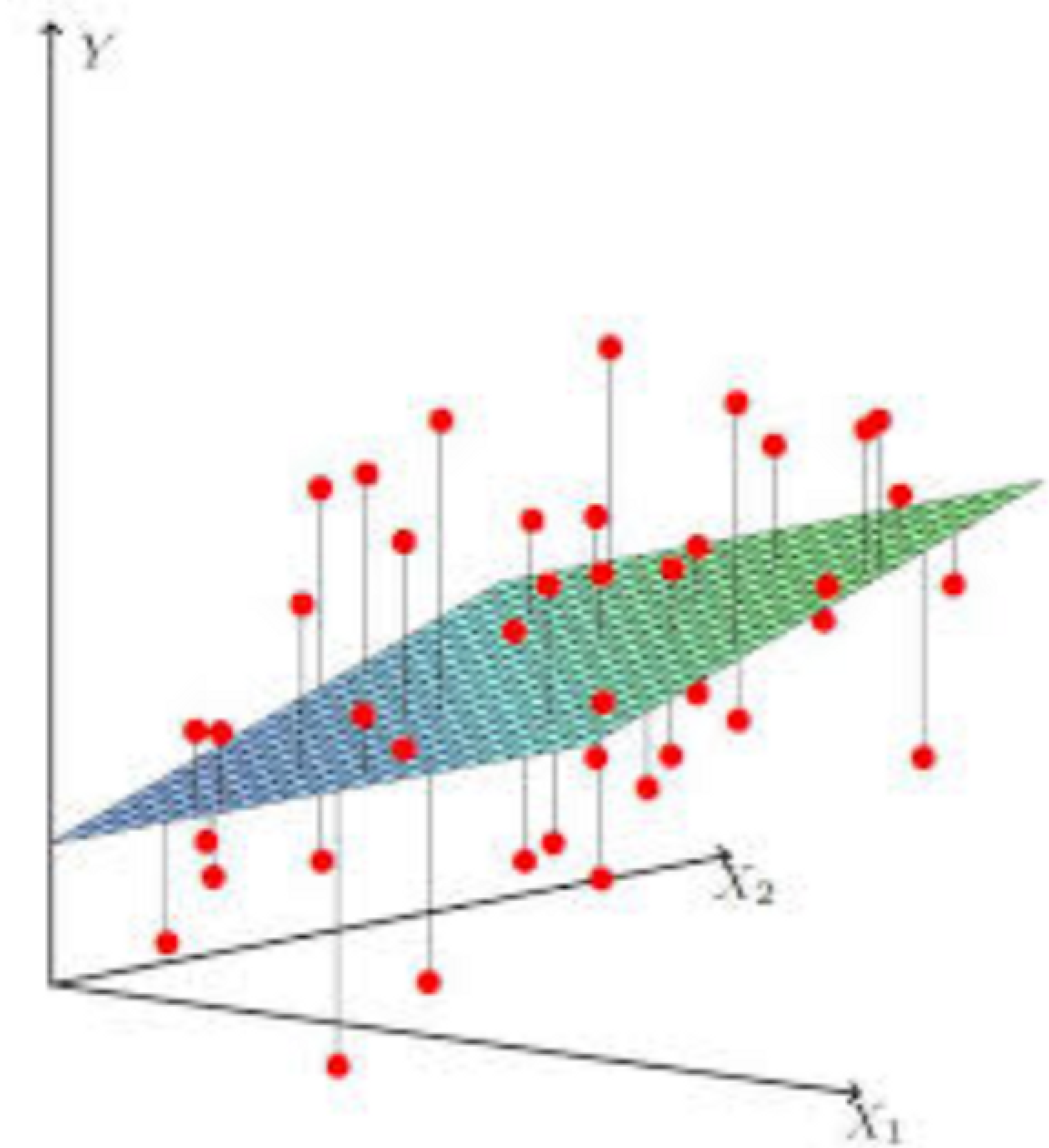
LET'S FOCUS ON MACHINE LEARNING

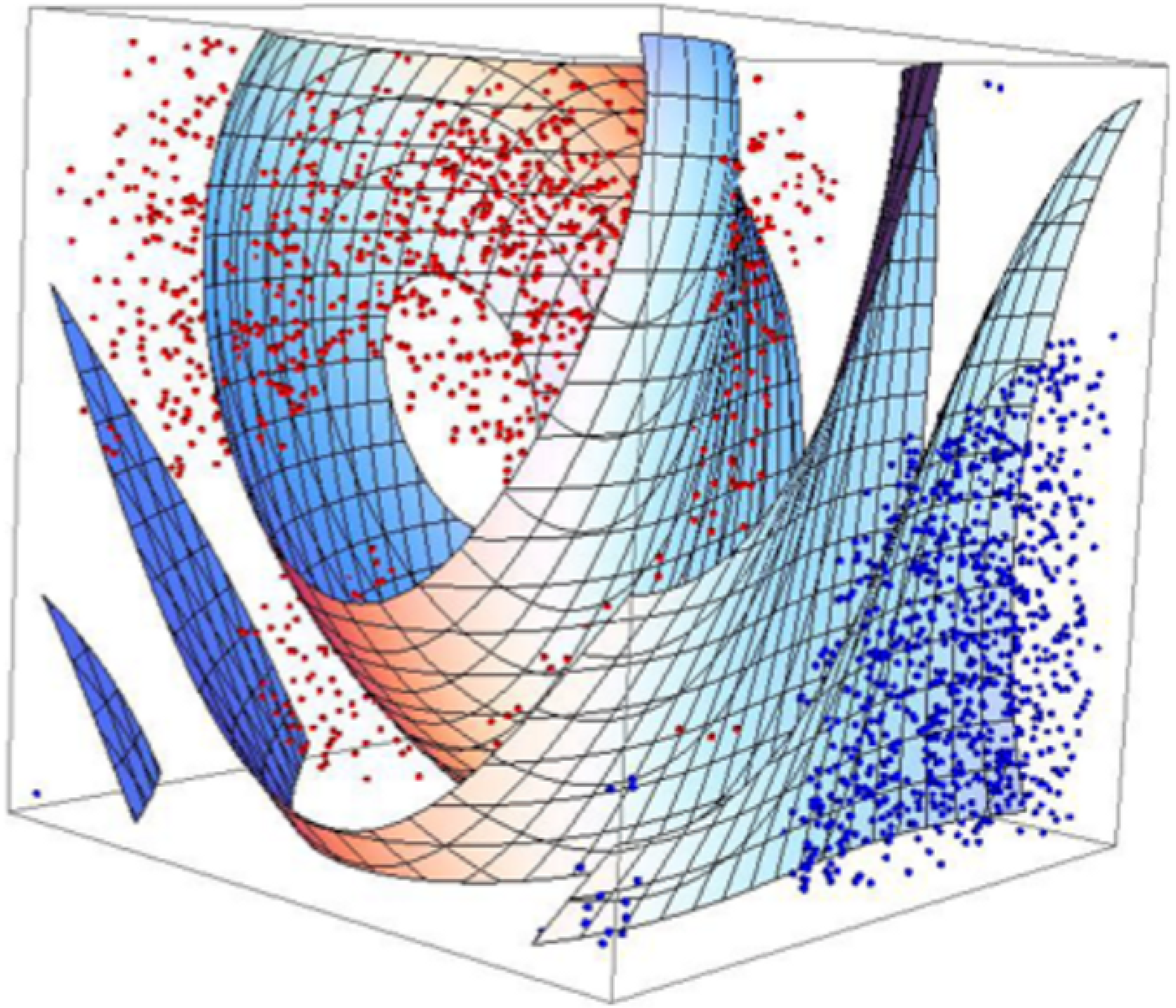
DATA -> INSIGHT ?

LET'S SOLVE THIS PROBLEM BY USING THE BIG DATA NONE OF US HAVE THE SLIGHTEST IDEA WHAT TO DO WITH
TOM FISH BURNE
© marketoonist.com

WHAT DOES ML DO?

DISCOVERS HARD-TO-FIND PATTERN IN DATA





GOAL OF ML

- ML attempts to 'learn' patterns in data with as little loss of information as possible

- ML learns from experience, not rules

DATA

- Historical Data: data collected over some period of time.

  Ex: Incomes and demographic qualities, zip-code, weather.

- Auto-Labelling Data: dataset built by using a tool or process.

  Ex: If we want a dataset of labeled animals, we can go to google images.

- Manual-Labelled Data: data labelled by a human analyst that provides the human class. This type of labelling is easier with crowd-sourcing.

  Ex: Detecting extremist language in different languages

DIFFERENT TYPES OF MACHINE LEARNING

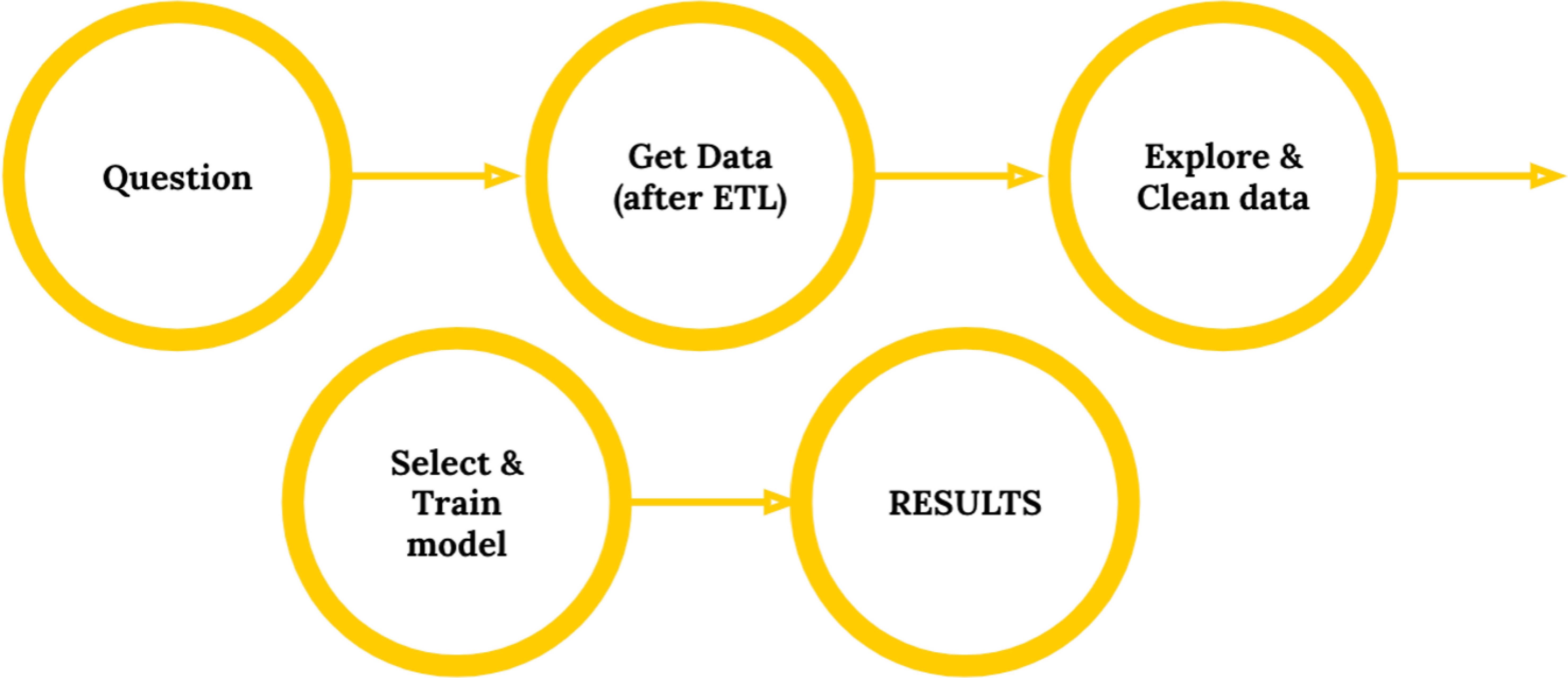
Supervised: The training data includes the outcome we want to know about.

Unsupervised: The training data does NOT include the outcome we want to know about

Reinforcement: The computer learns through trial and error

SUPERVISED LEARNING

WHAT TO DO



Question

Get Data (after ETL)

Explore & Clean data

Select & Train model

RESULTS

WHAT'S ETL

- Extract: reading data from database (different types and sources)

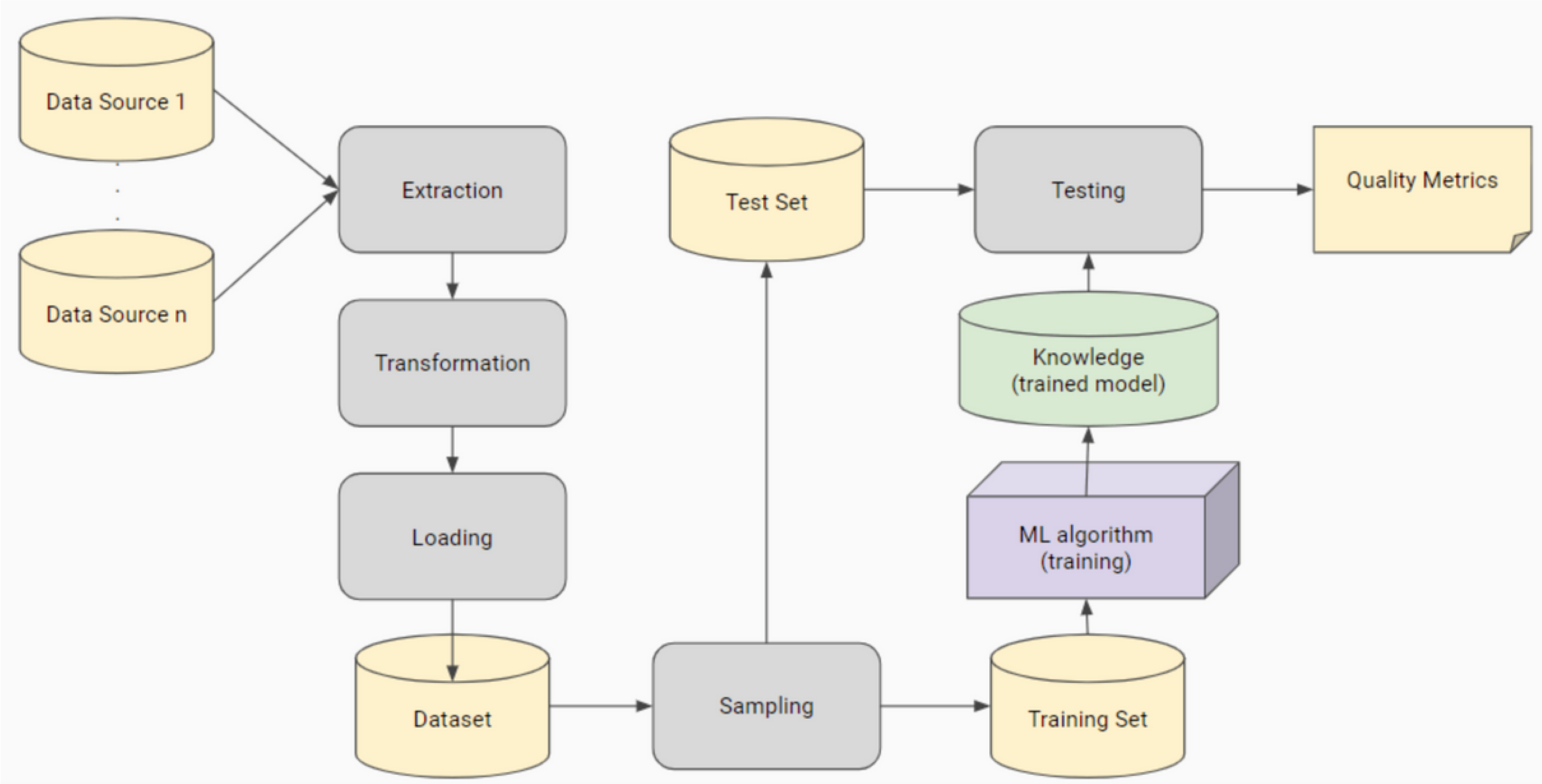
- Transform: convert from previous form to format database needs.

- Load: writing data to database

(what Data Engineers do)

SUPERVISED LEARNING

MAIN IDEA



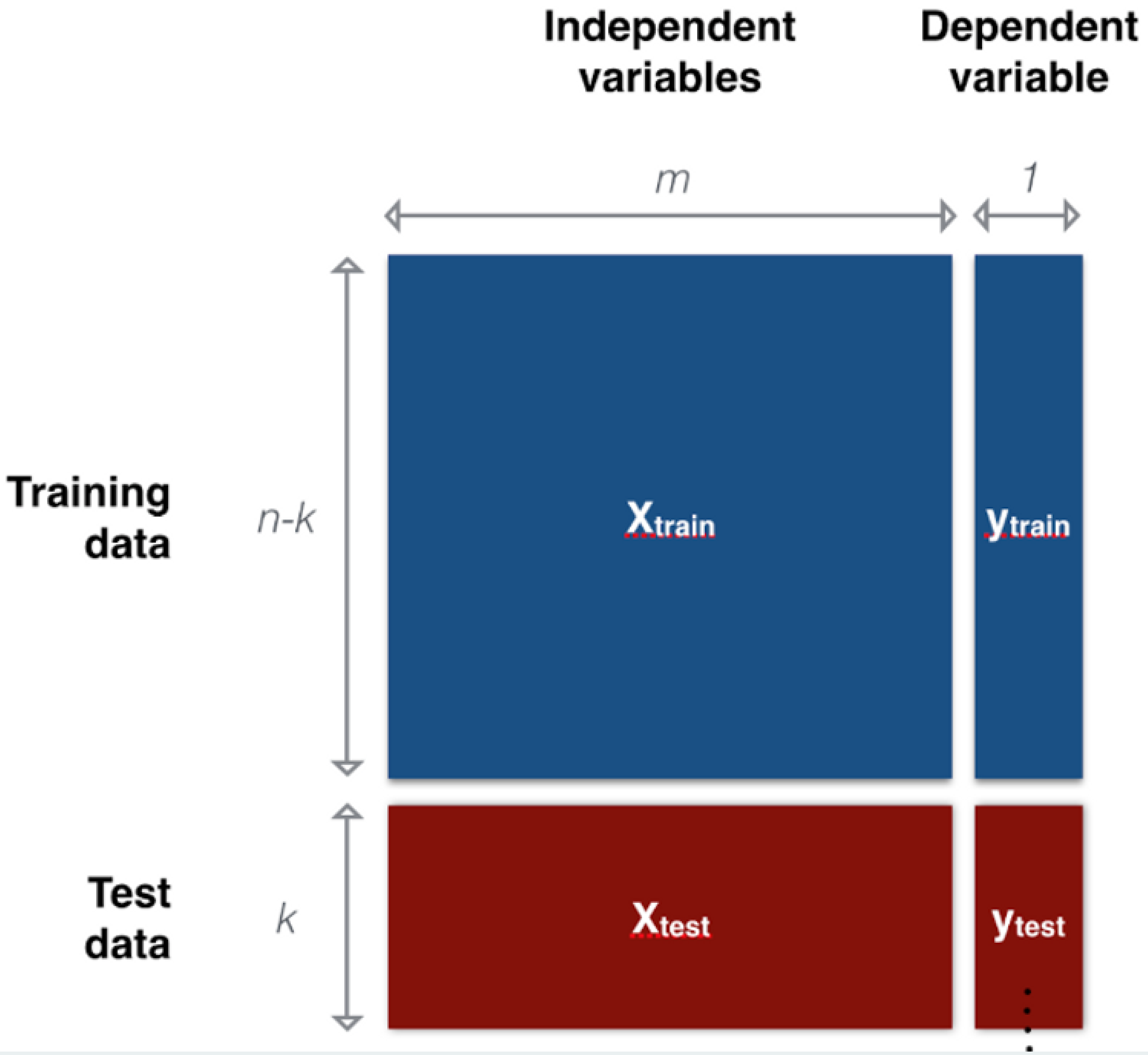
SUPERVISED LEARNING

MAIN IDEA

Train Set: 60%-80% of all the data available. Used to have the model learn.

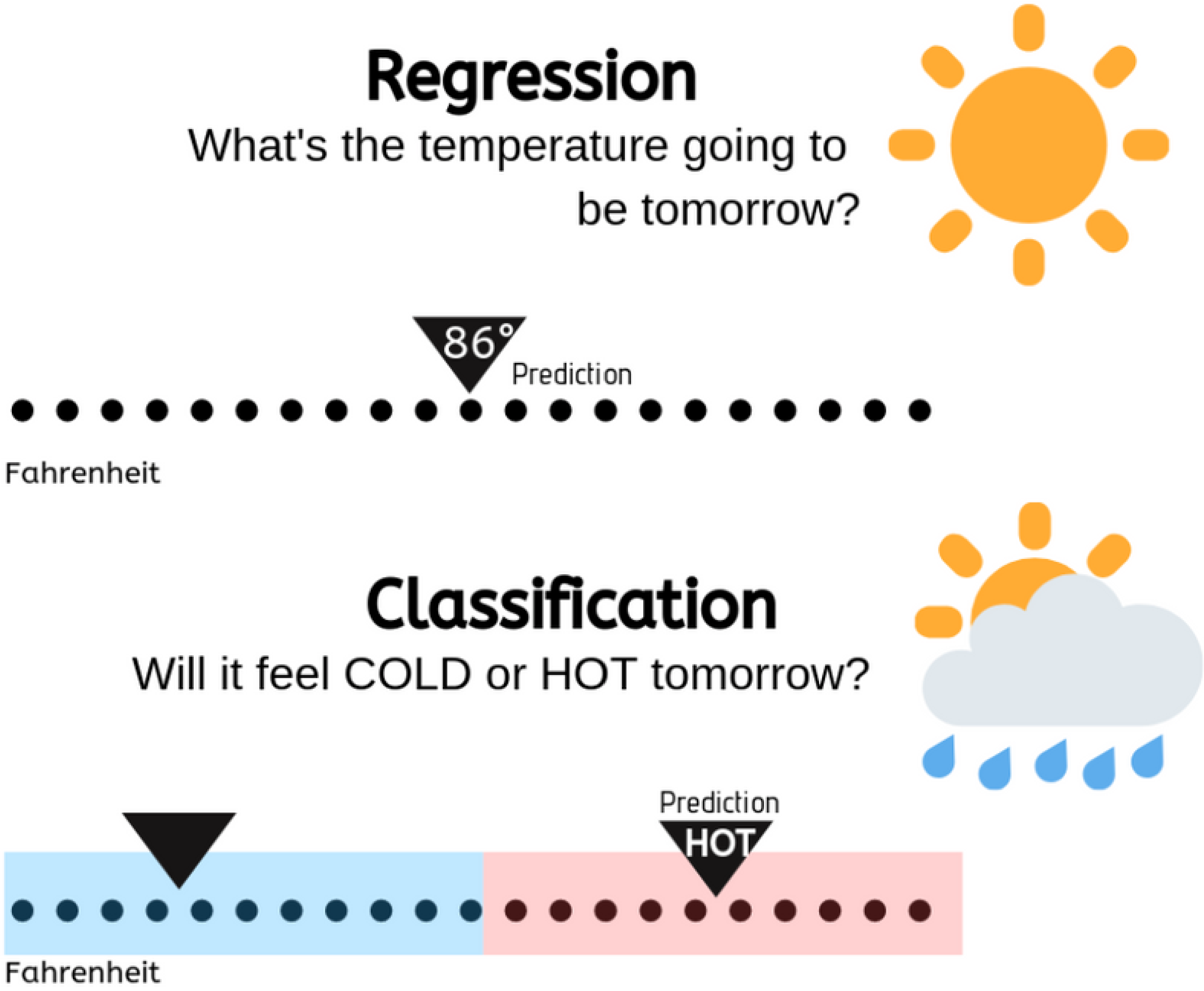
Test Set: Used with the train algorithm to extract predicted values.

Evaluation: Comparison between predicted values and real ones to determine how well it makes predictions.



SUPERVISED LEARNING

CLASSIFICATION VS REGRESSION



Regression

What's the temperature going to be tomorrow?

Classification

Will it feel COLD or HOT tomorrow?

DEALING WITH FEATURES

WHAT TO DO BEFORE CREATING THE MODEL

- Conversion: same units

- Feature scaling: make all the inputs on the same scale

- Missing values: find a way to manage data that is not in the dataset.
  Why is it missing? How?

- Categorical data: algorithms handle numbers. One-Hot-Encoding!!
  What happens when we have categories in our data?

ANY QUESTIONS ?