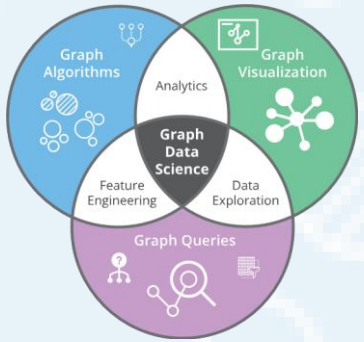


Habib University
shaping futures

CS343 Graph Data Science

Graph Algorithms to Enhance Machine Learning

Chapter #8, Graph Algorithms, Mark

Muhammad Qasim Pasta

qasim.pasta@sse.habib.edu.pk

Machine Learning

- Machine learning is a branch of artificial intelligence that enables computers to **learn from data** without being explicitly programmed
- Machine learning algorithms use **mathematical methods** to find patterns and make predictions or classifications based on data

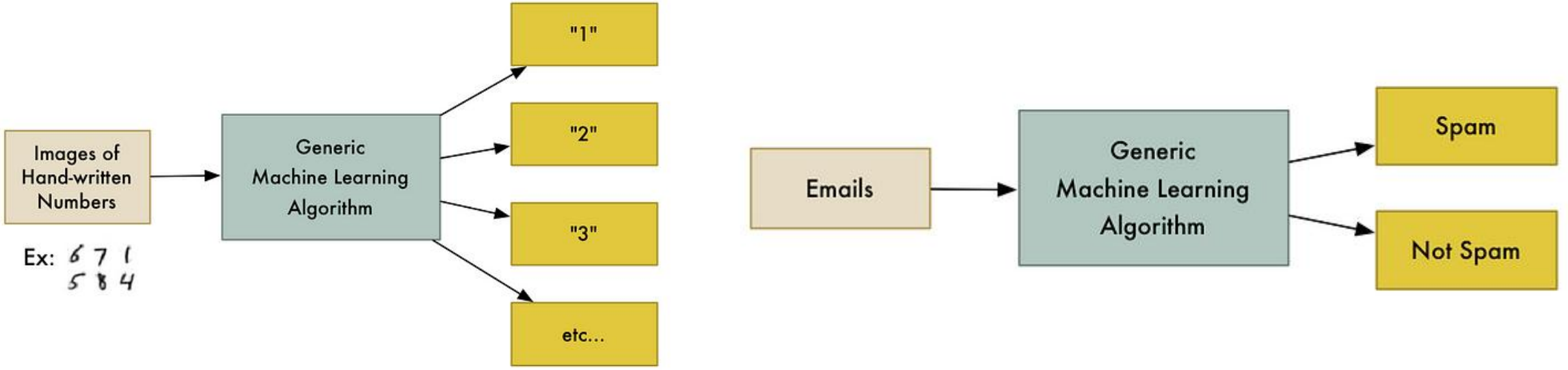


Image Credit: Adam Geitgey

Types of Machine Learning

- Supervised
- Unsupervised
- Reinforcement
- Self-supervised
- Semi-supervised

Supervised Learning

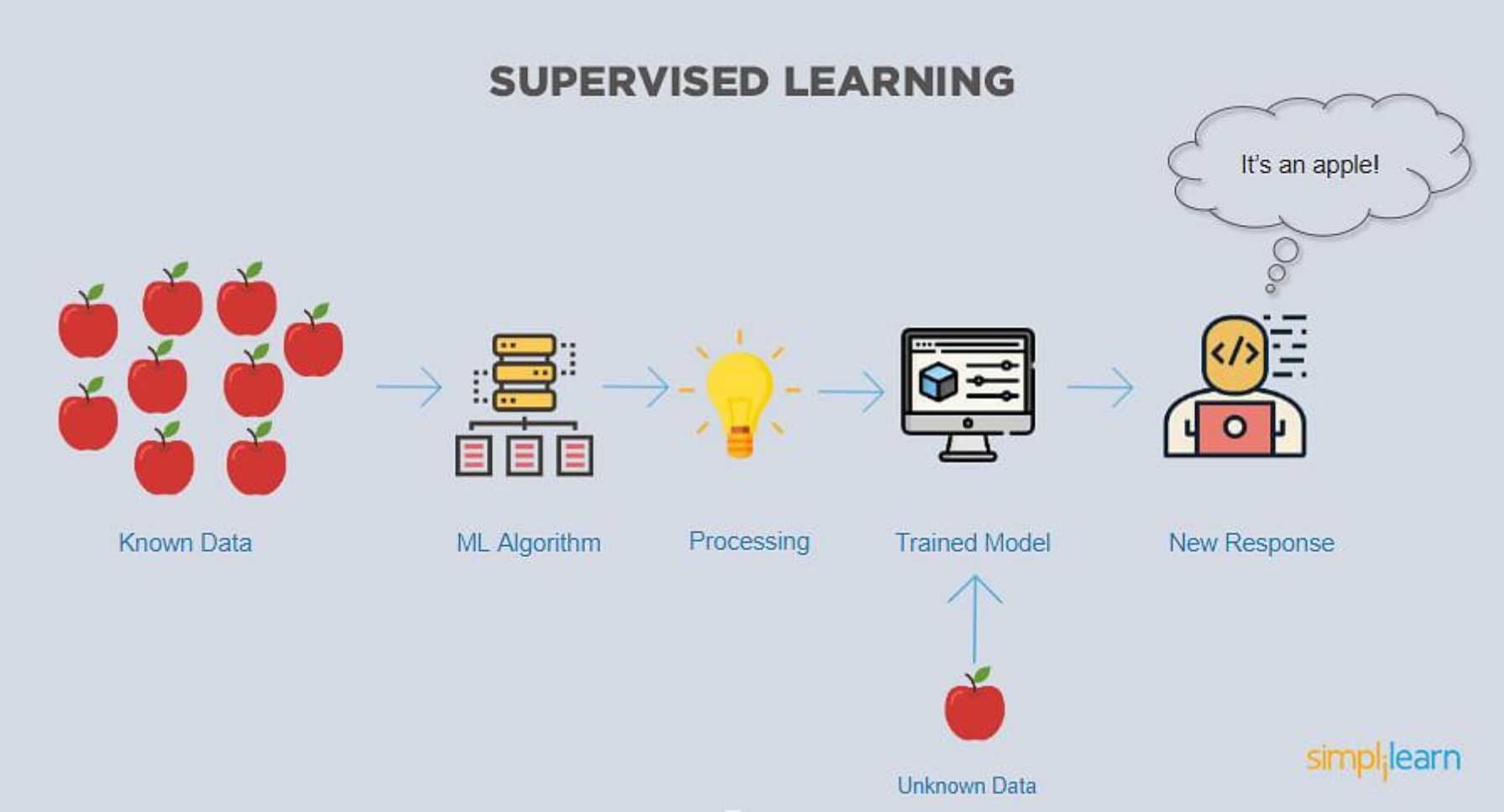


Image Credit: Simple Learn

Unsupervised Learning

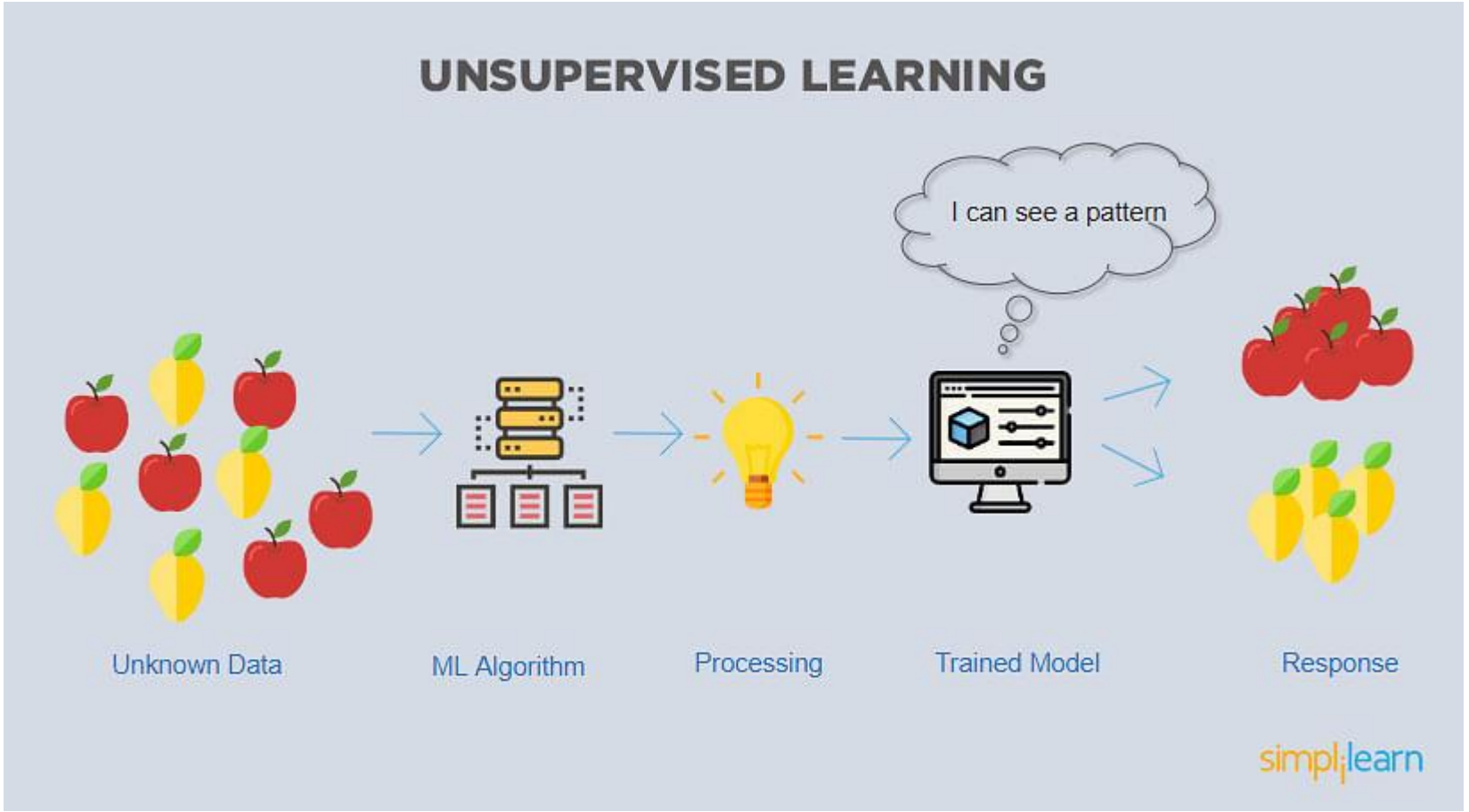


Image Credit: Simple Learn

Reinforcement Learning

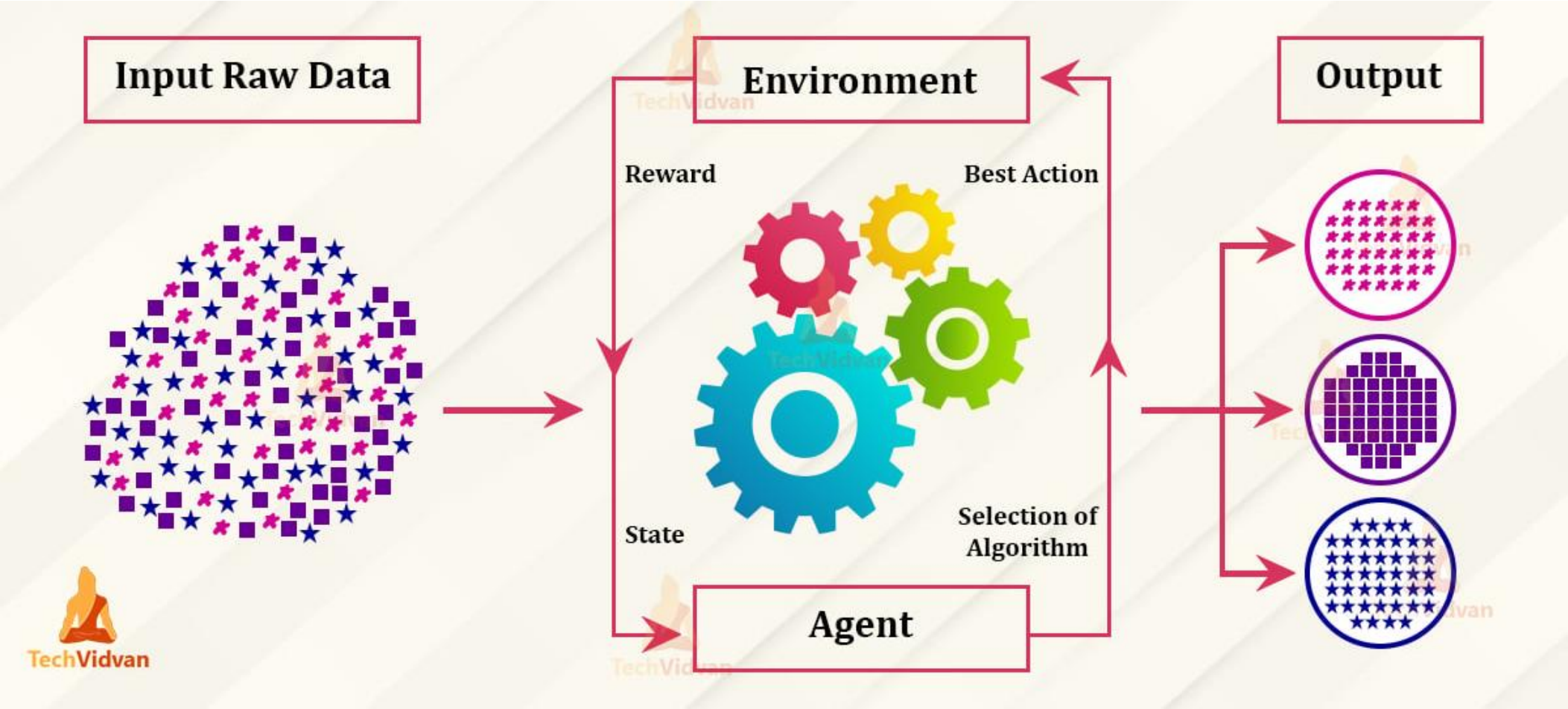


Image Credit: Tech Vidan

Machine Learning Pipeline

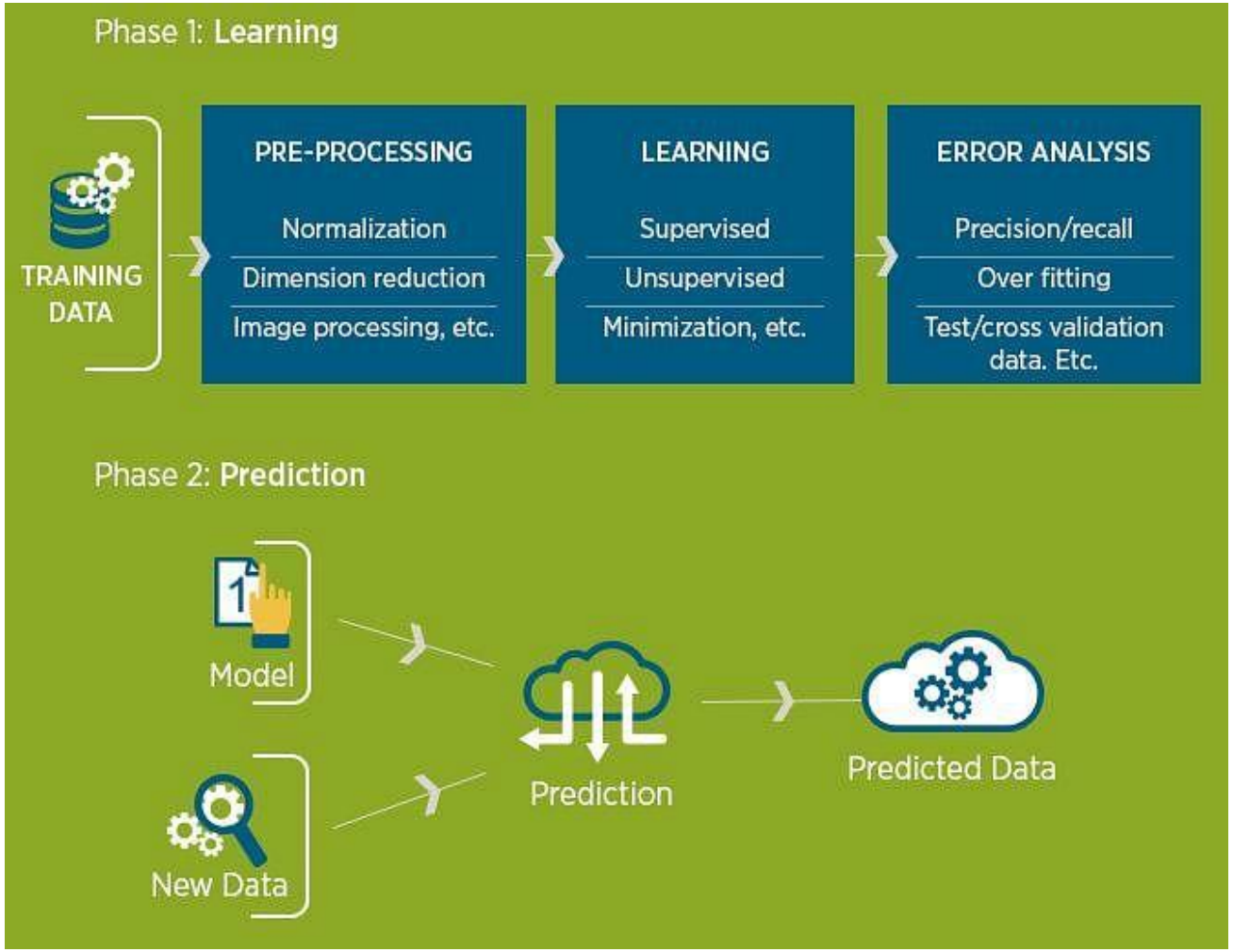


Image Credit: Simple Learn

Feature Extraction

- Transforming raw data into a format that is suitable for analysis and model building.
- Simplifies complex data by extracting meaningful patterns.
- Reduces dimensionality, making the data more manageable and improving computational efficiency.
- Enhances model performance by focusing on important characteristics of the data.

- **Feature selection** is the process of determining the subset of extracted features that are most important or influential to a target goal.
  - Selects the most relevant features based on criteria such as correlation, significance, or importance scores.

Connected Features

- Features extracted from the structure of the data.
- Graph-local queries based on parts of the graph surrounding a node
- Graph-global queries that use graph algorithms to identify predictive elements

- Graph for feature selection?
  - What if map features to nodes in a graph, create relationships based on similar features, and then compute the centrality of features.

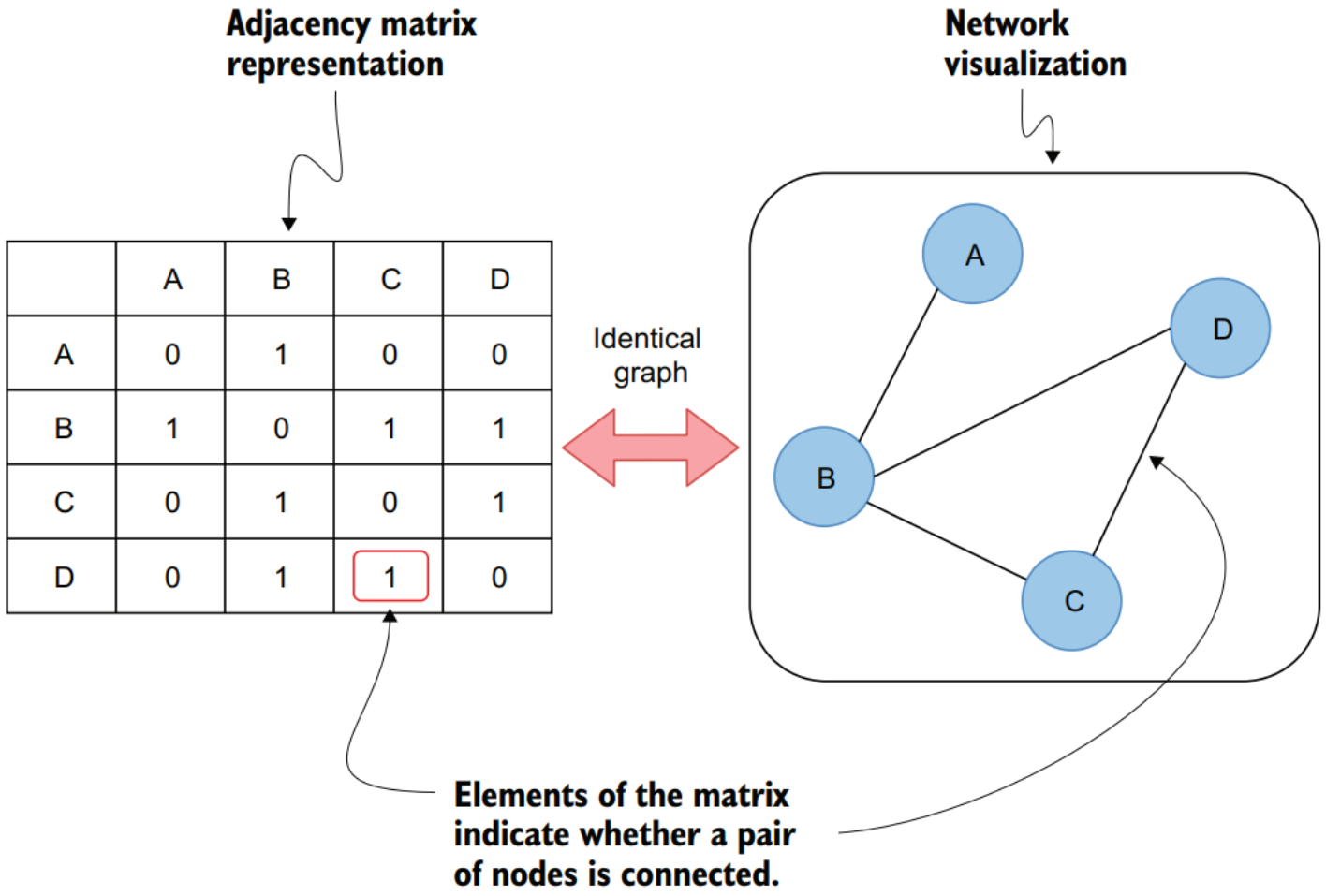
Node Embedding

- Maps each node in a graph to a low-dimensional vector while preserving structural relationships.

Figure 9.1 Body mass index chart

Adjacency Matrix

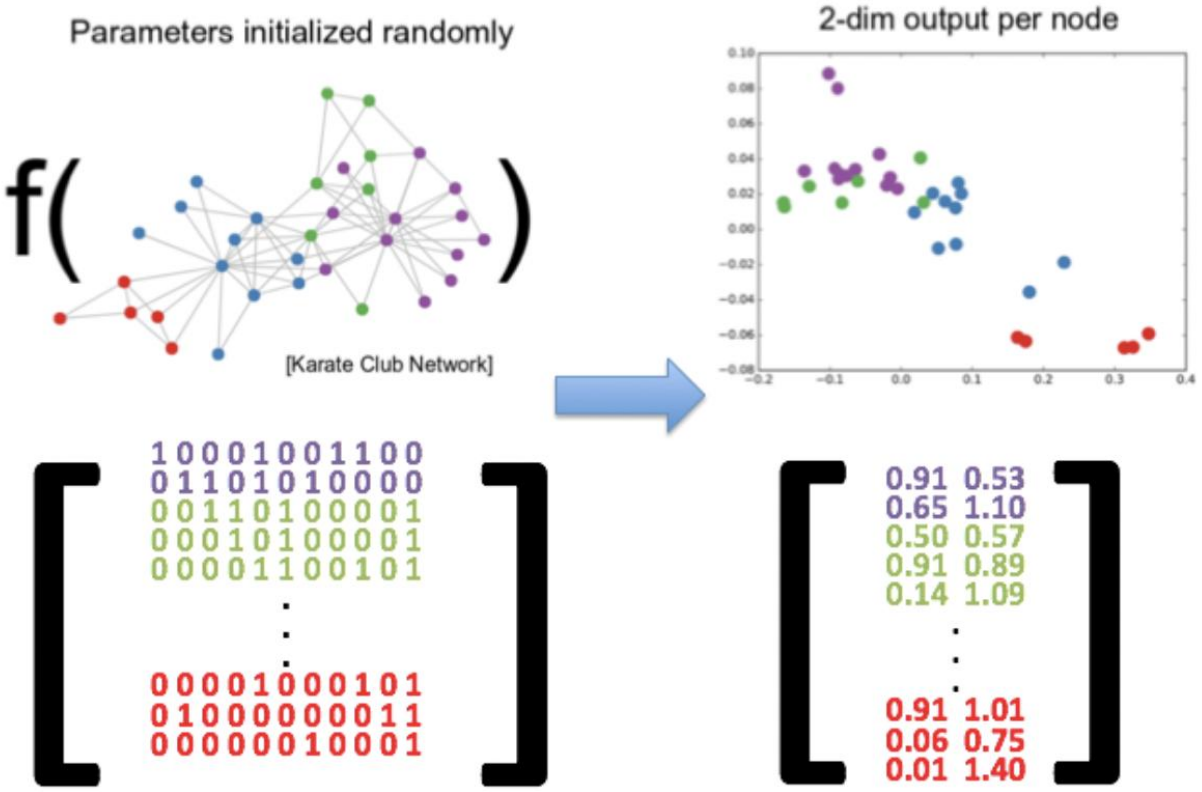
- High-dimensional network representation



	A	B	C	D
A	0	1	0	0
B	1	0	1	1
C	0	1	0	1
D	0	1	1	0

Node Embedding

- Lower-dimensional vectors can be used in a downstream machine learning workflow, or they can be used to infer a similarity network using the nearest neighbor graph algorithm.



Homophily vs Structure Holes

FastRP

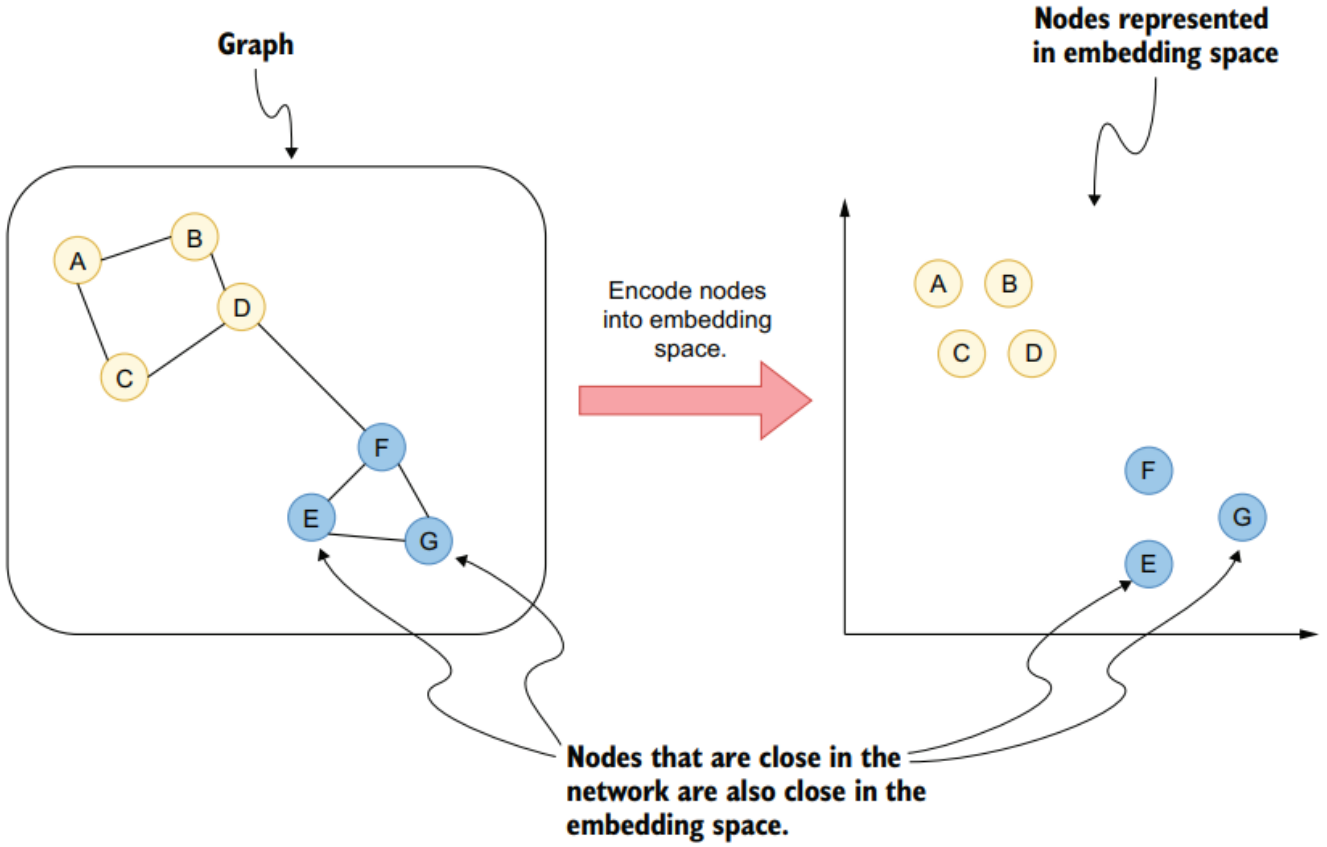


Figure 9.4 Homophily approach to node embedding

Homophily vs Structure Holes

node2vec

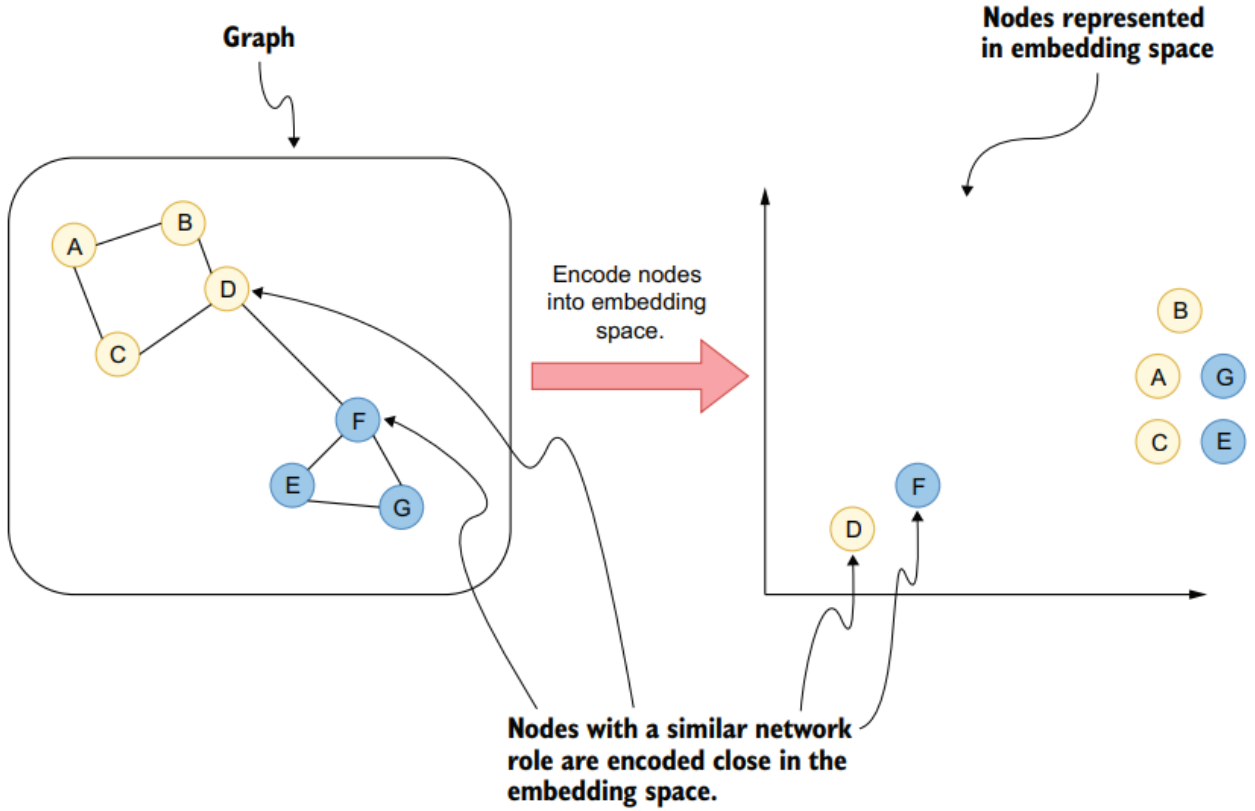


Figure 9.5 Structural roles approach to node embedding

Transductive vs Inductive

- Transductive:
  - cannot calculate embeddings for nodes not seen during the initial embedding calculation
  - cannot simply retrieve the embeddings for the new unseen nodes
  - New node: must calculate the embeddings for the whole graph
  - must also retrain the classification model.
- Inductive
  - can calculate embeddings for unseen nodes during the initial computation
  - can calculate the embedding for the new node without recalculating embeddings for the whole graph
  - don’t need to retrain the classification model for every new node.

Node Classification Example

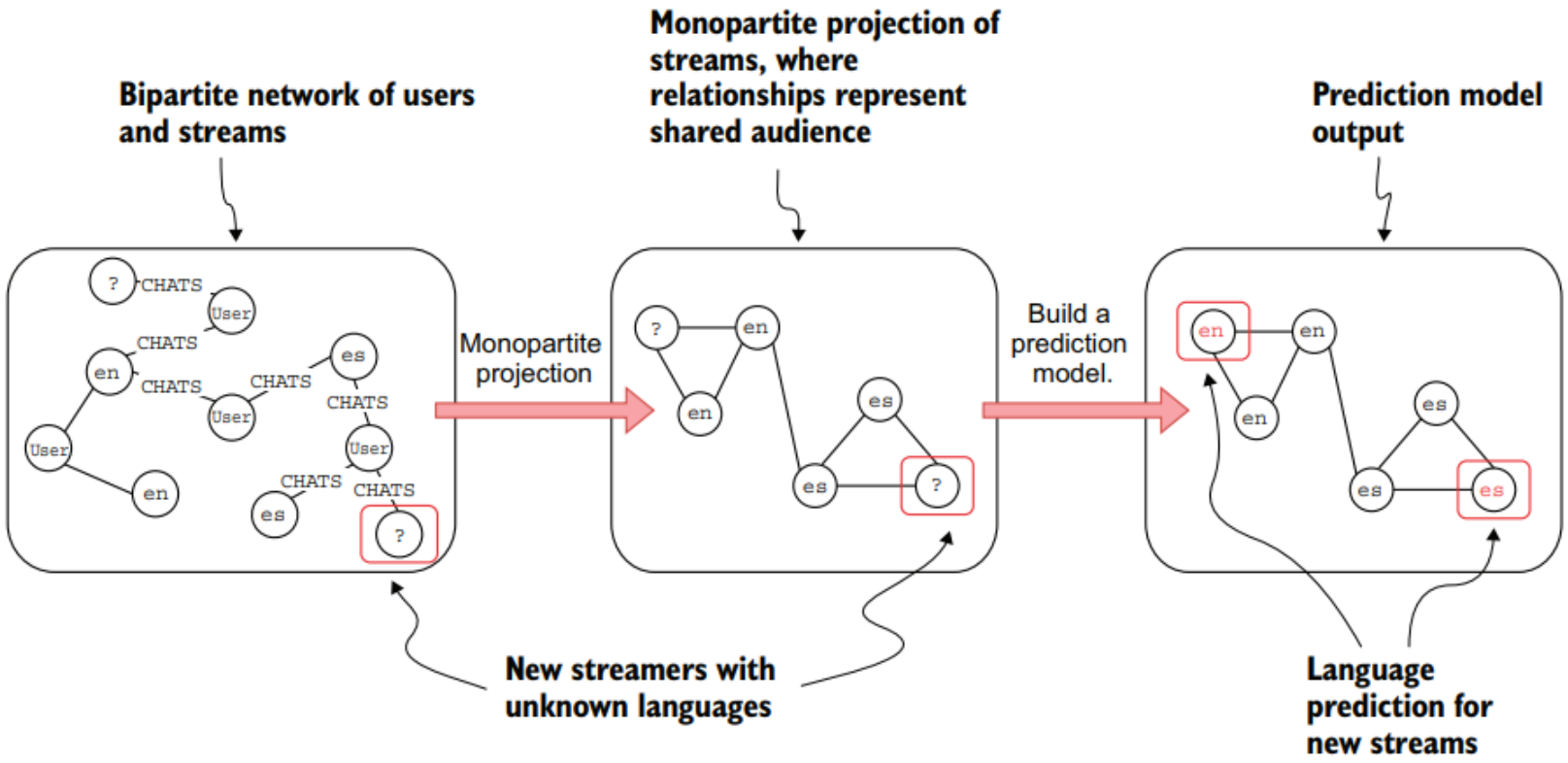


Figure 9.6 The process of predicting the language for new streams