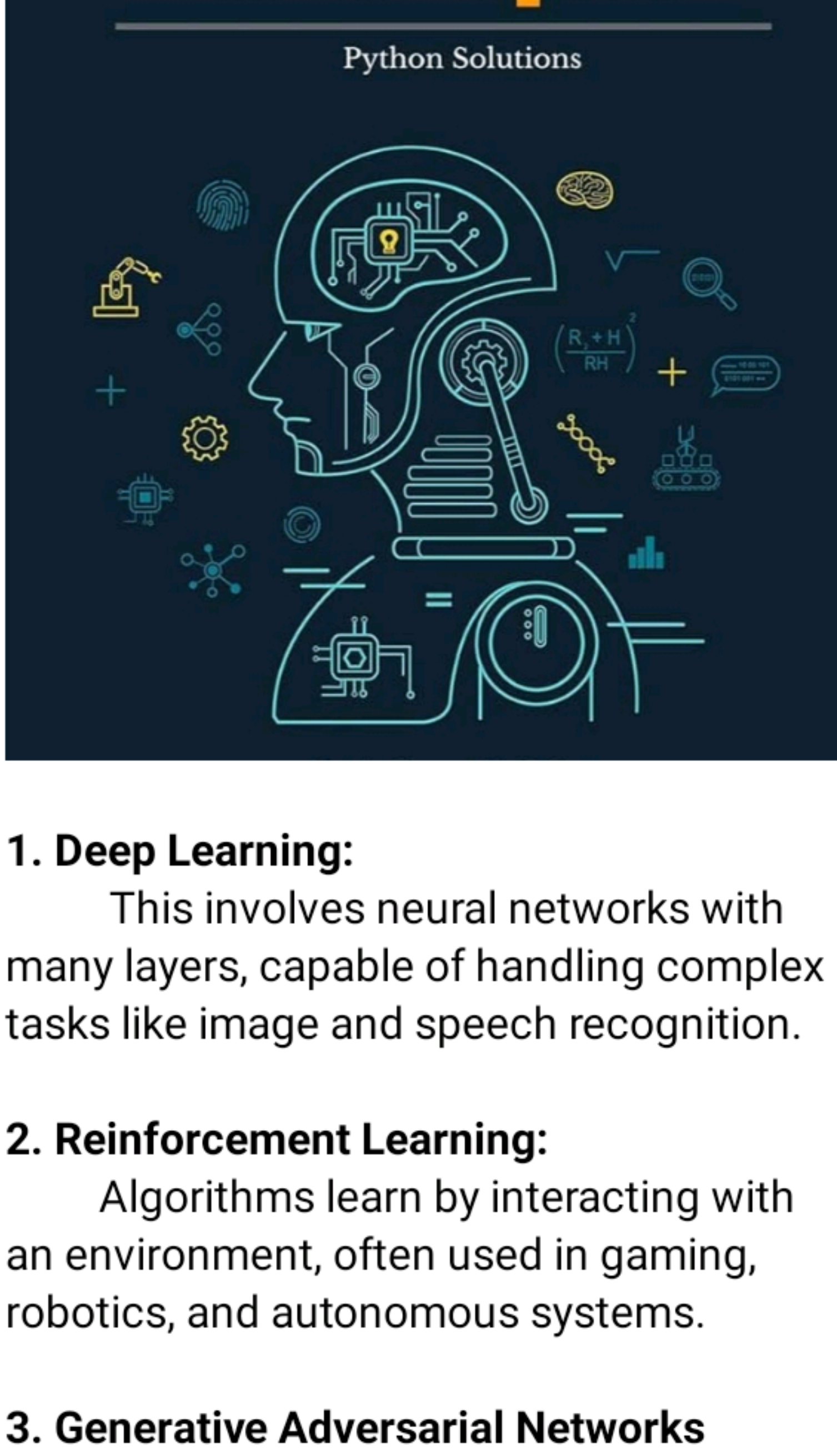


## DATA ANALYTICS AND COGNOS:

## Phase -2

## Machine learning

# Advanced Machine Learning



Used for ge  
resembles real da

in image generation and data augmentation.

**4. Transfer Learning:**  
Leveraging pre-trained models on

reduce training time and data requirements.

### 5. Natural Language Processing (NLP):

Techniques for understanding and generating human language, used in chatbots, translation, and sentiment

## 6. Ensemble Learning:

Combining multiple models to improve overall performance, often

random forests, gradient boosting, and stacking.

## 7. AutoML:

and platforms that support the development process

accessible to non-experts.

**8. Explainable AI (XAI):**  
Techniques that provide insights into how AI models make decisions,

## 9. Time Series Analysis

Specialized methods for handling data with temporal dependencies, used in forecasting and trend analysis.

## 10. Anomaly Detection:

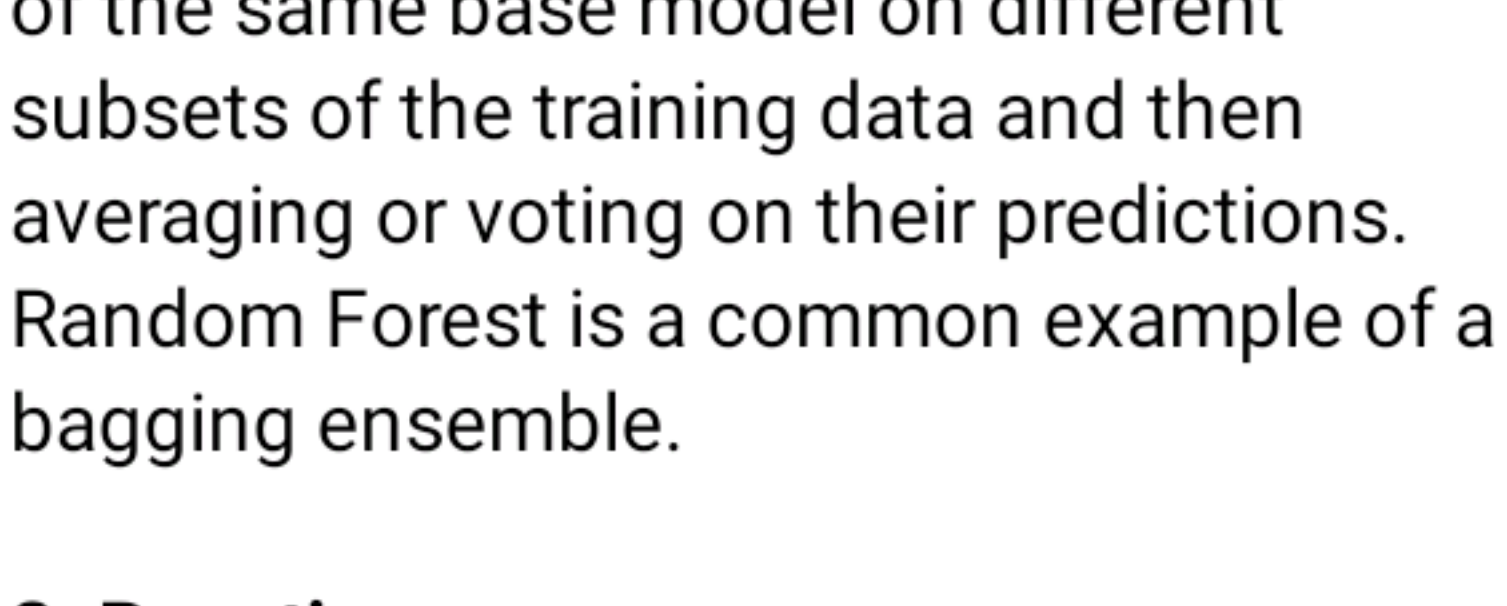
outliers in data, important  
detection and system m

## Ensemble models

A diagram illustrating data processing. On the left, a stack of four horizontal cylinders represents data. Two arrows branch out from this stack to the right. The top arrow points to a gear icon labeled "Algorithm 2". The bottom arrow points to a gear icon labeled "Algorithm 3". To the right of these algorithms, there are two horizontal blue lines. The top line has a small blue circle at its right end. The bottom line has a small blue circle at its right end. A vertical dashed orange line is positioned between the algorithms and the blue lines.

### 1. Bagging (Bootstrap Aggregating):

Involves training multiple instance

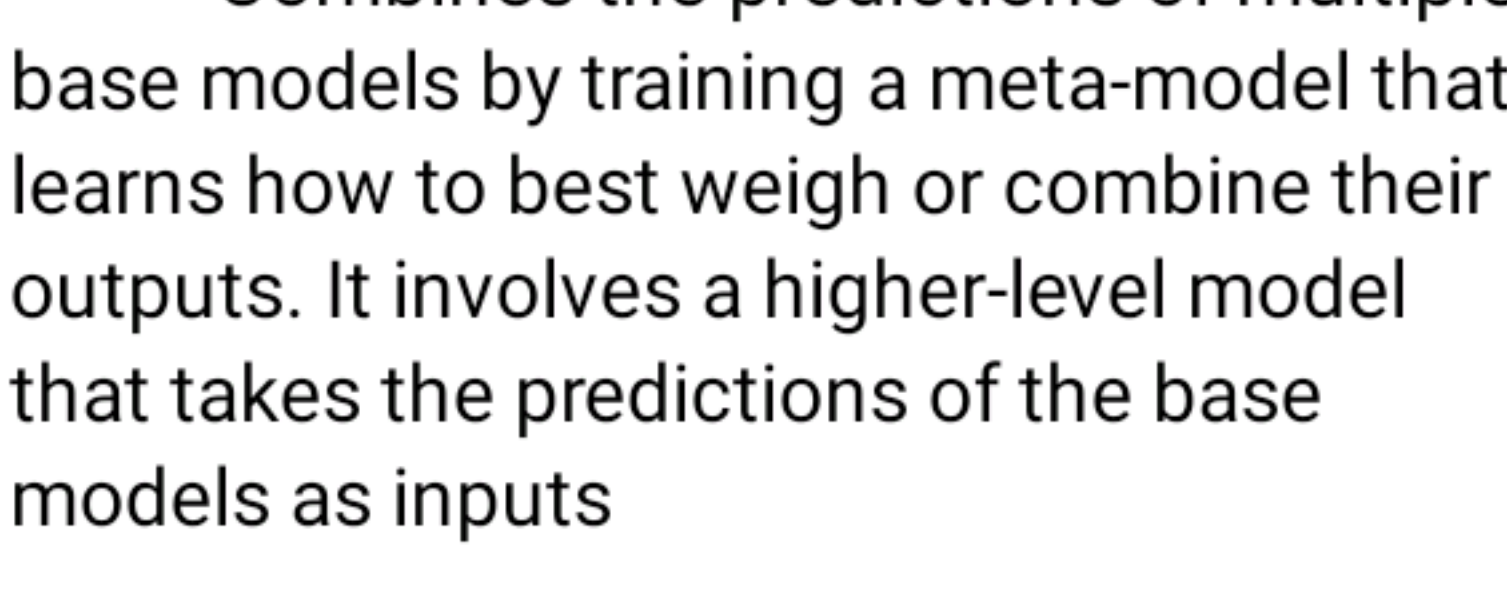


Builds a sequence of base models  $\{M_1, M_2, \dots, M_n\}$  where each subsequent model focuses on the residual error of the previous model.

correcting the errors of the previous ones. AdaBoost and Gradient Boosting are widely used boosting algorithms.

Algorithm 1

### 3. Stacking:



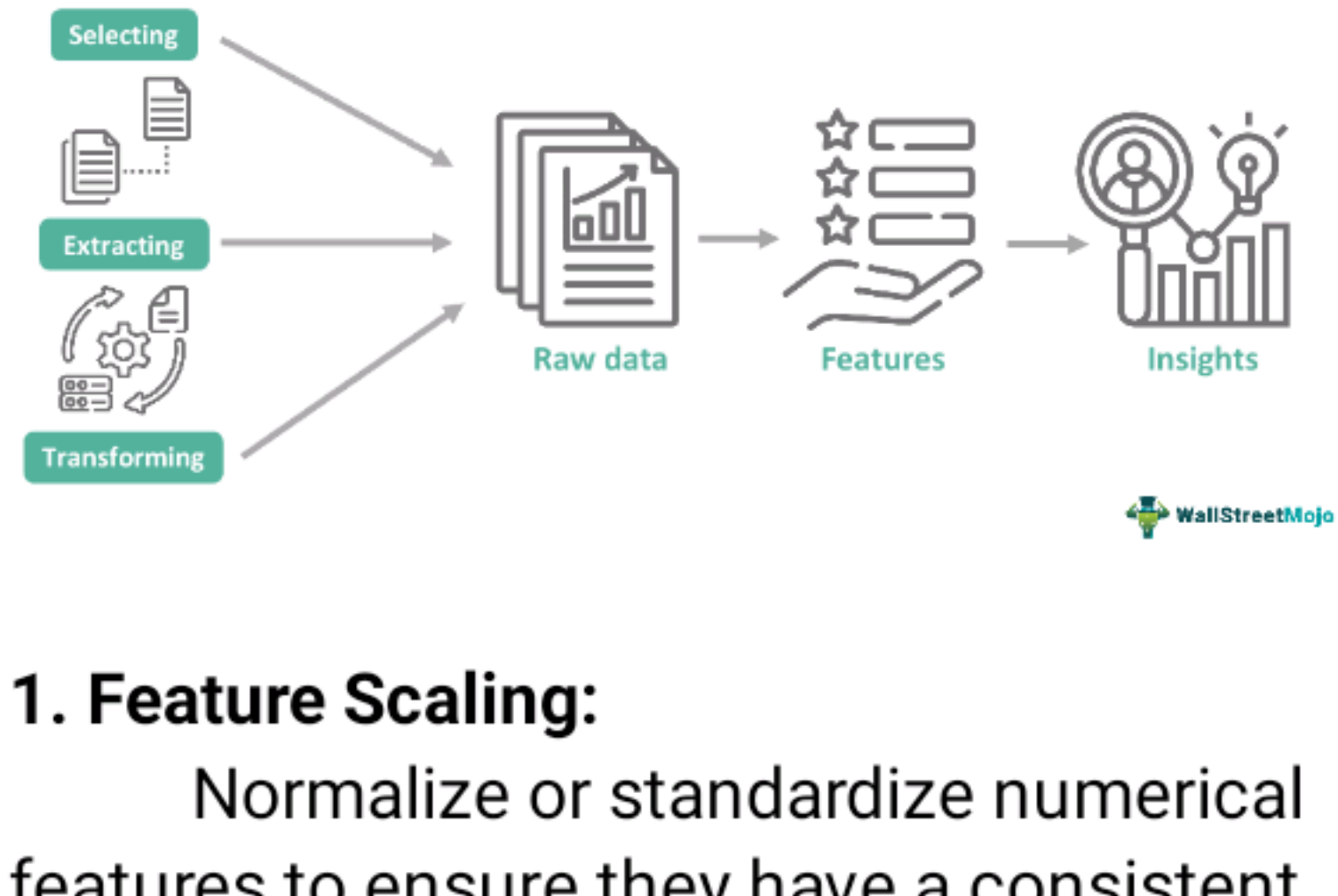
#### 4. Randomly

Similar

bootstrapping

bootstrapping the data, it randomly selects a subset of features for each base model. This is often used with high-dimensional data

## What is Feature Engineering?



scale. This is important  
sensitive to feature  
K-Means

## 2. One-Hot Encoding:

Convert categorical variables into binary vectors to make them suitable for machine learning algorithms that require numerical input.

### 3. Feature Extraction

If you have high-dimensional data, consider techniques like Principal Component Analysis (PCA) or t-Distributed Stochastic Neighbor Embedding (t-SNE)

important information

#### 4. Binning or Discretization:

Convert continuous variables into discrete bins to capture non-linear relationships and reduce noise in the data.

Create new features by combining two or more existing features:

help capture interactions between variables.

**6. Time-Based Features:**  
If working with time series data

generate features like lag averages, or time-based a

## 7. Text Feature Engineering:

For NLP tasks, preprocess text data by tokenizing, stemming, or using techniques like TF-IDF or Word