

# MACHINE LEARNING

## Introduction to Machine Learning

Muhammad Afif Hendrawan, S.Kom., M.T.



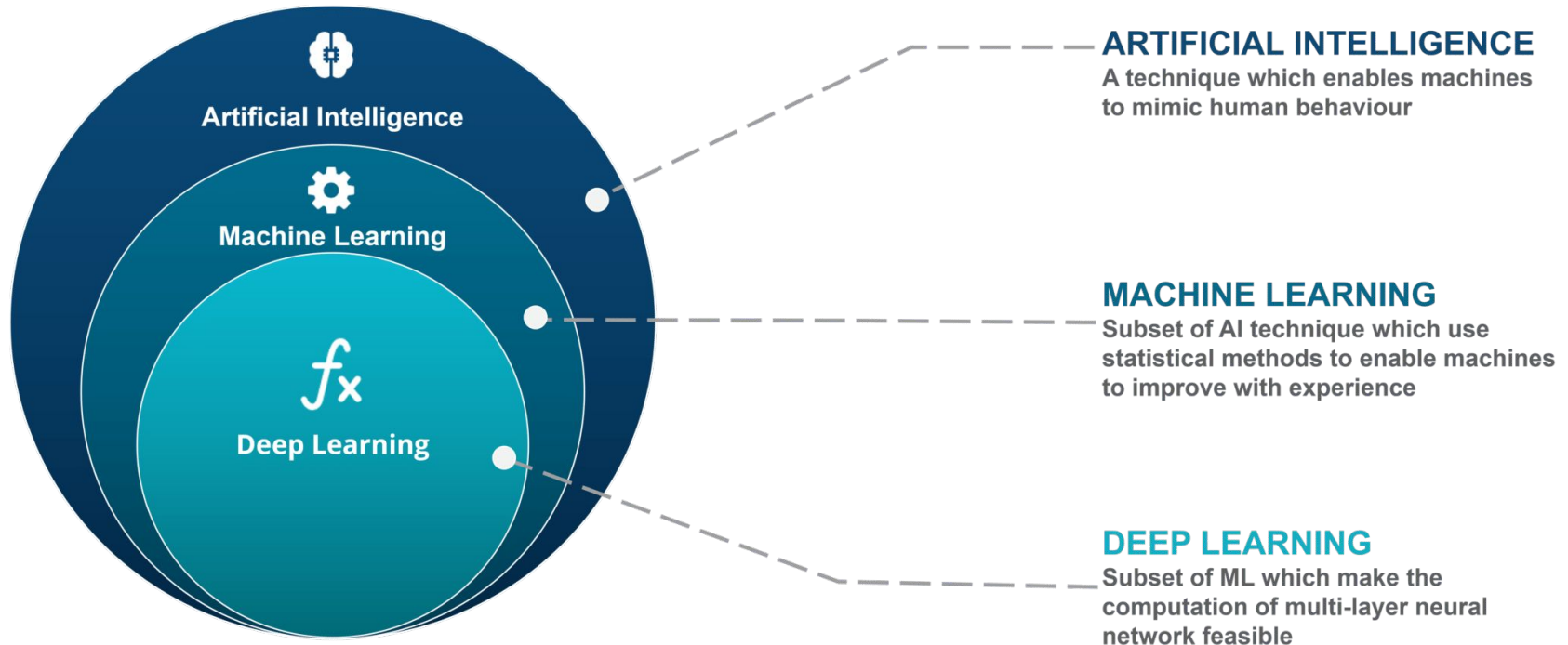
# Outlines

- The relation between artificial intelligence and machine learning
- What is machine learning?
- Machine learning implementation
- Types of machine learning
- Data types on machine learning context
- Training, testing, and validation data
- Introduction to machine learning model evaluation
- Setup machine learning development environment



# The Relation Between AI and ML

# AI vs. ML. vs. DL



<https://www.edureka.co/blog/ai-vs-machine-learning-vs-deep-learning/>



# What is Machine Learning

# Formal Definition

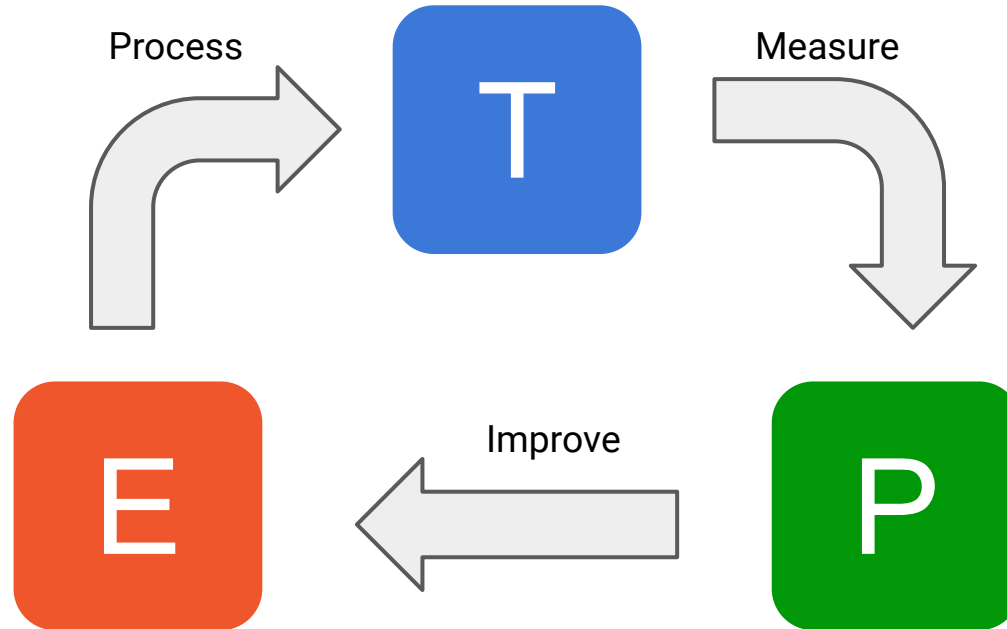
- “Computational methods using **experience** to improve performance or to make accurate predictions” – Mohri, et al., 2012
- “Machine learning is the design and study of software artifacts that **use past experience** to inform future decisions; machine learning is the study of programs that **learn from data**” – Hackeling, 2017

"A program can be said to learn from experience 'E' with respect to some class of tasks 'T' and performance measure 'P', if its performance at tasks in 'T', as measured by 'P', improves with experience 'E'."



Tom Mitchell

# Tom Mitchell's ML Concept

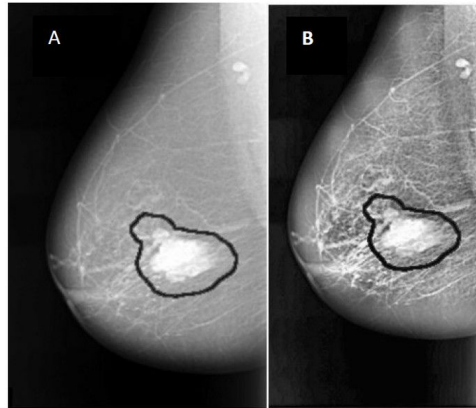
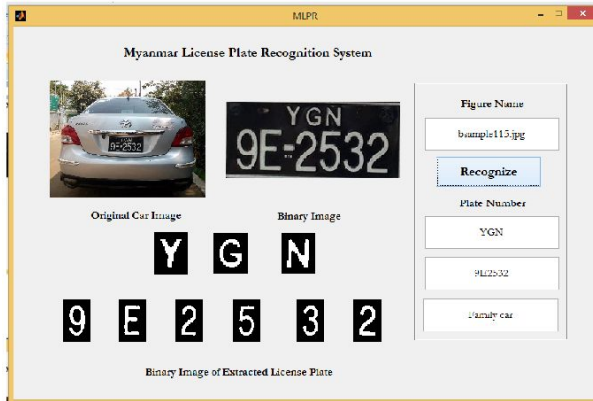
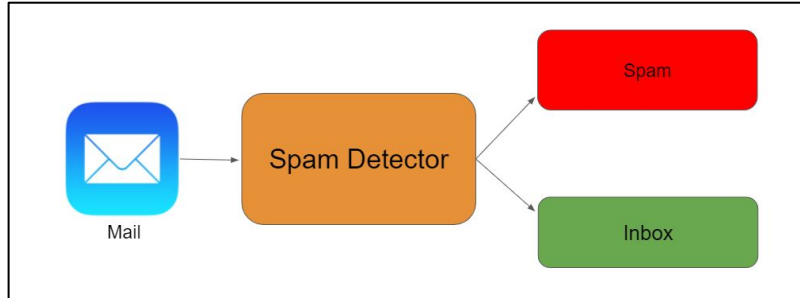





# Machine Learning Implementation





# ML in Real World #1



**Added to cart** ×

 **FastAPI - The Complete Course 2023 (Beginner + Advanced)**  
Eric Roby Go to cart

**Frequently Bought Together**

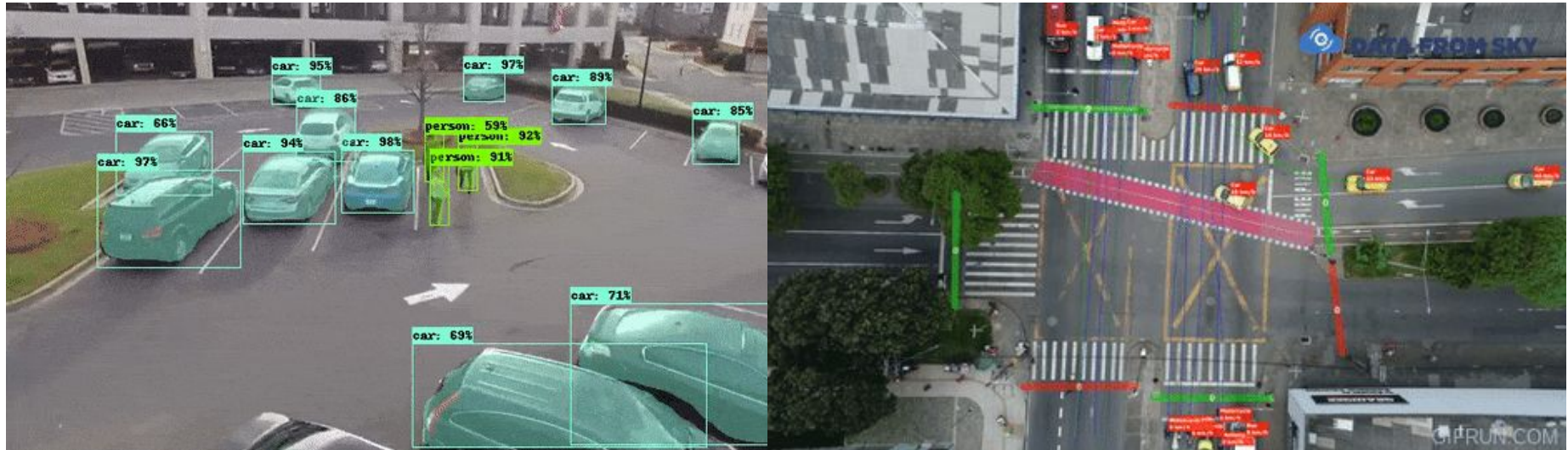
	<b>FastAPI - The Complete Course 2023 (Beginner + Advanced)</b> Eric Roby 4.6 ★★★★★ (2,282) Bestseller	<b>Rp109,000</b> <del>Rp649,000</del>
	<b>Complete FastAPI masterclass from scratch</b> Catalin Stefan 4.7 ★★★★★ (892)	<b>Rp99,000</b> <del>Rp549,000</del>

**Total: Rp208,000** ~~Rp1,198,000~~ Add all to cart

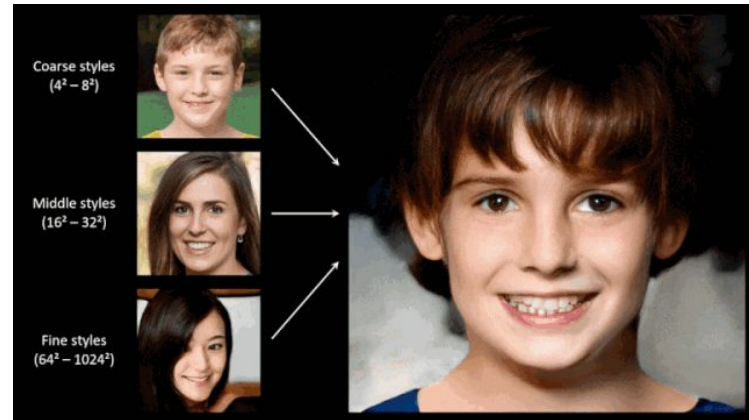
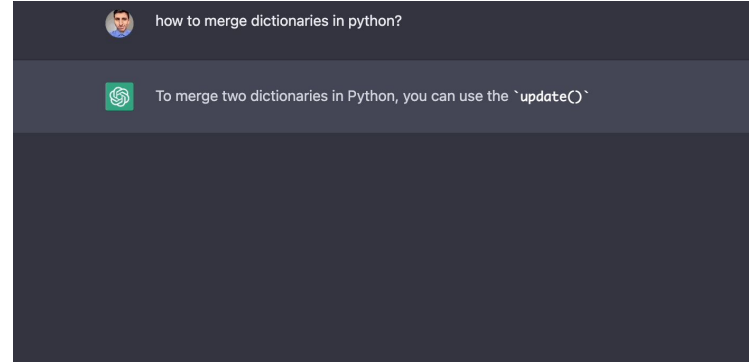
**Related topics**

- API
- SQLite
- Python Network Programming
- Java Game Development
- Flask

## ML in Real World #2

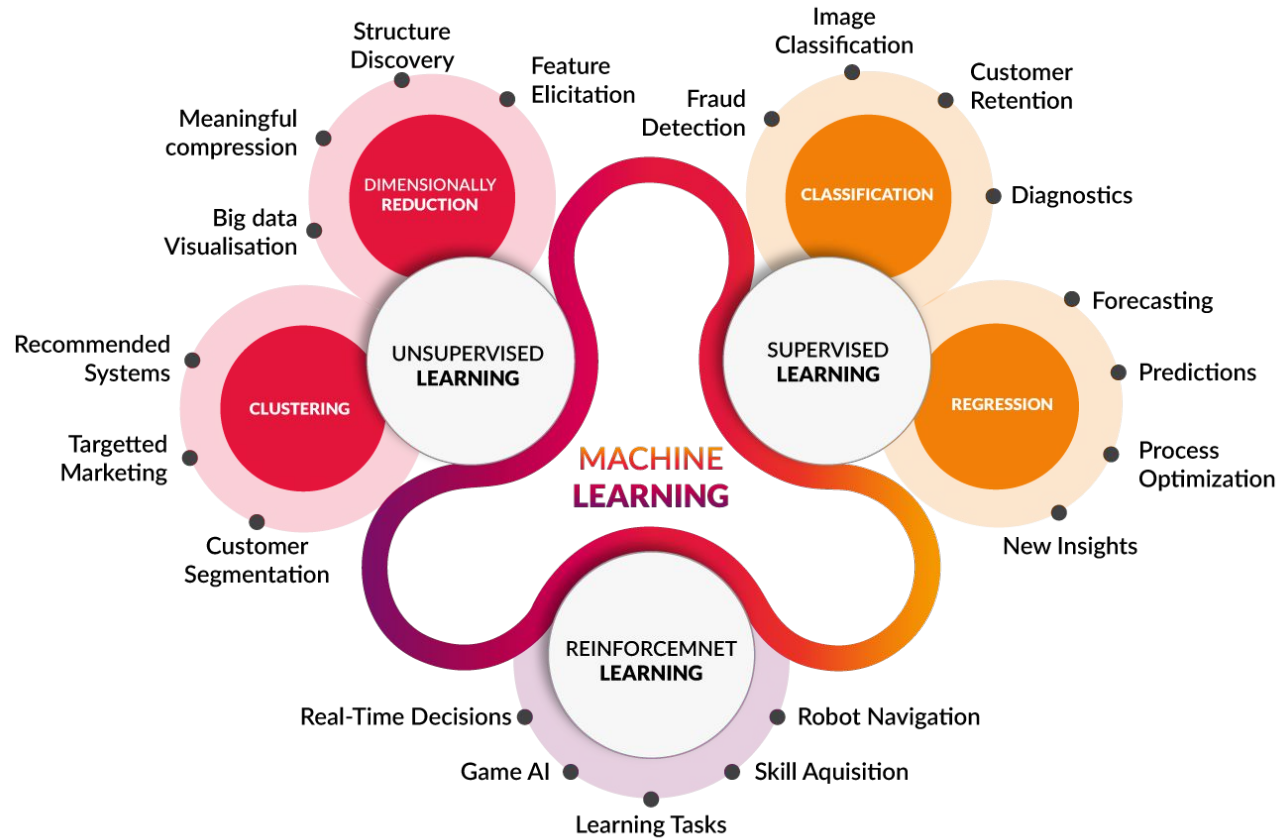


# ML in Real World #3

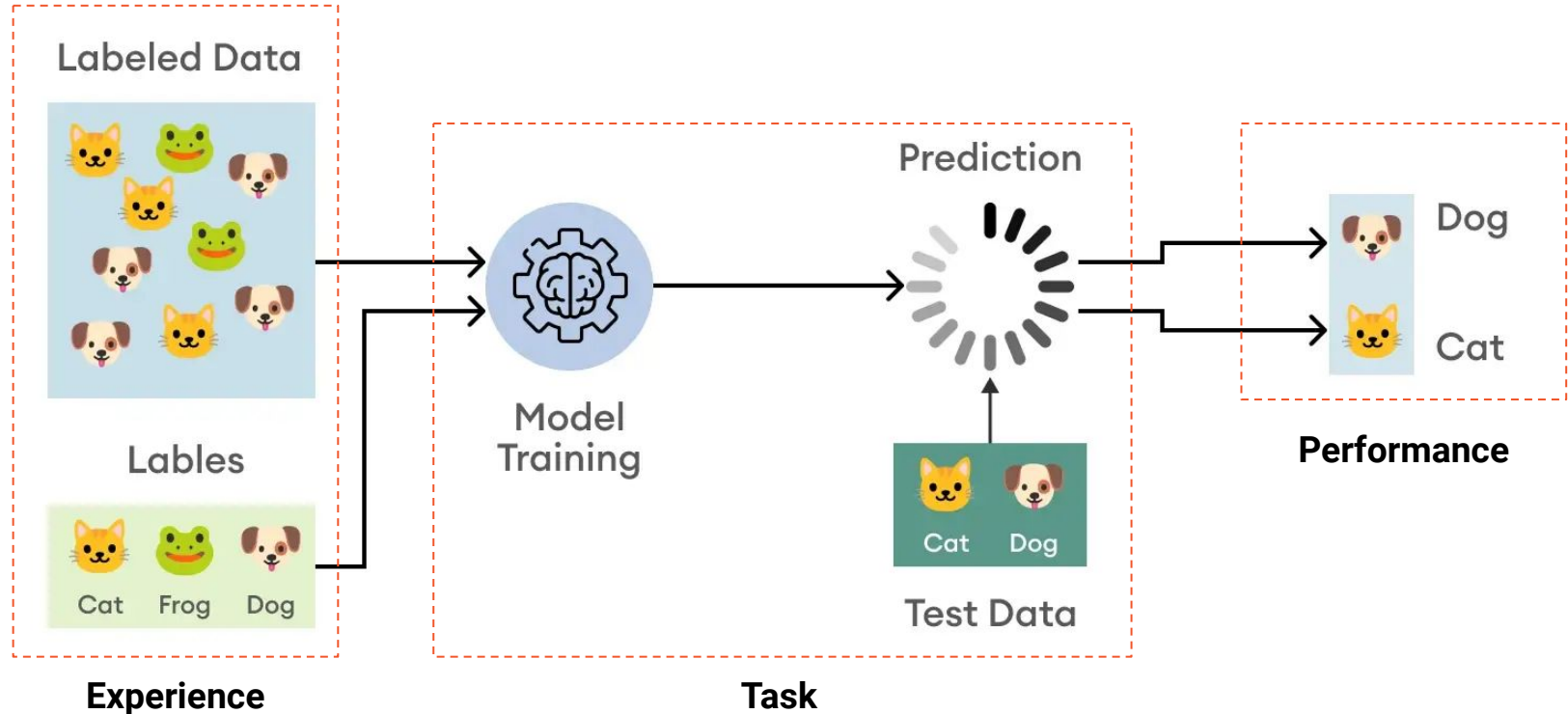




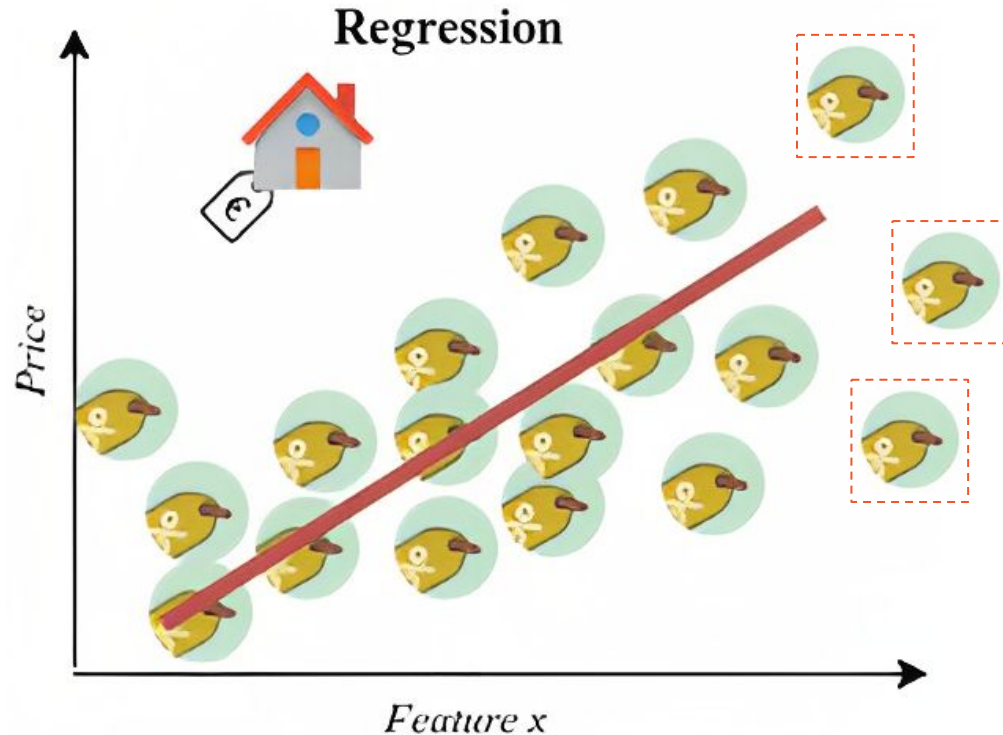
# Types of Machine Learning



# The Concept of Supervised Learning - Classification Tasks

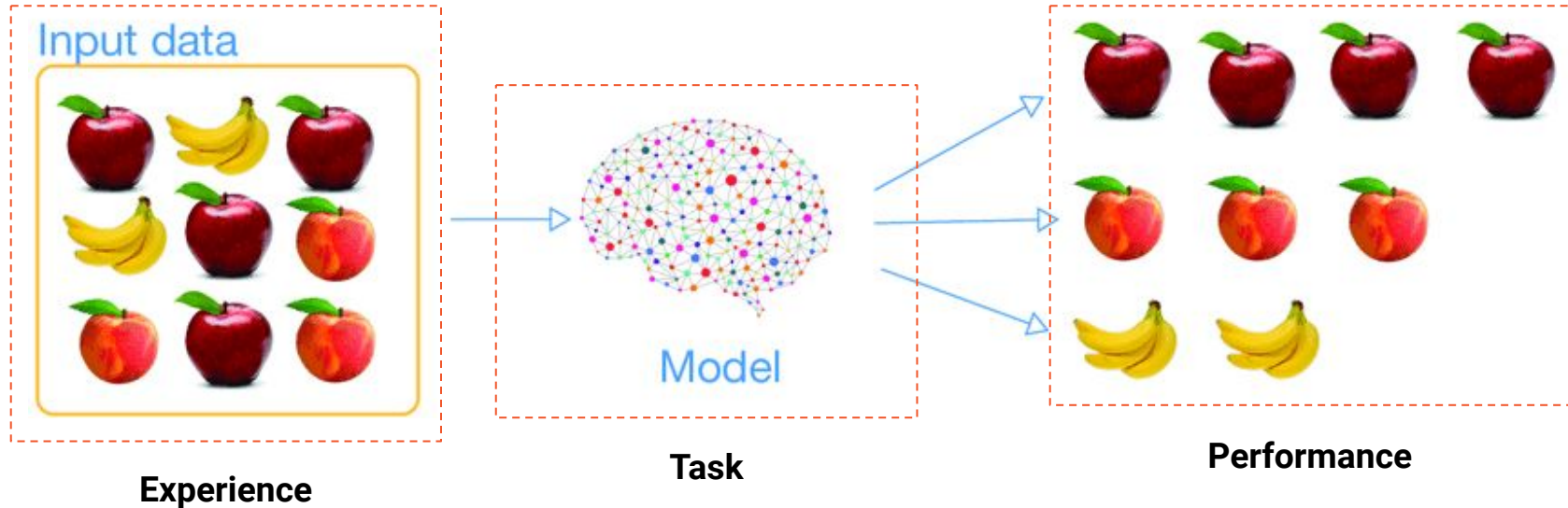


# The Concept of Supervised Learning - Regression Tasks



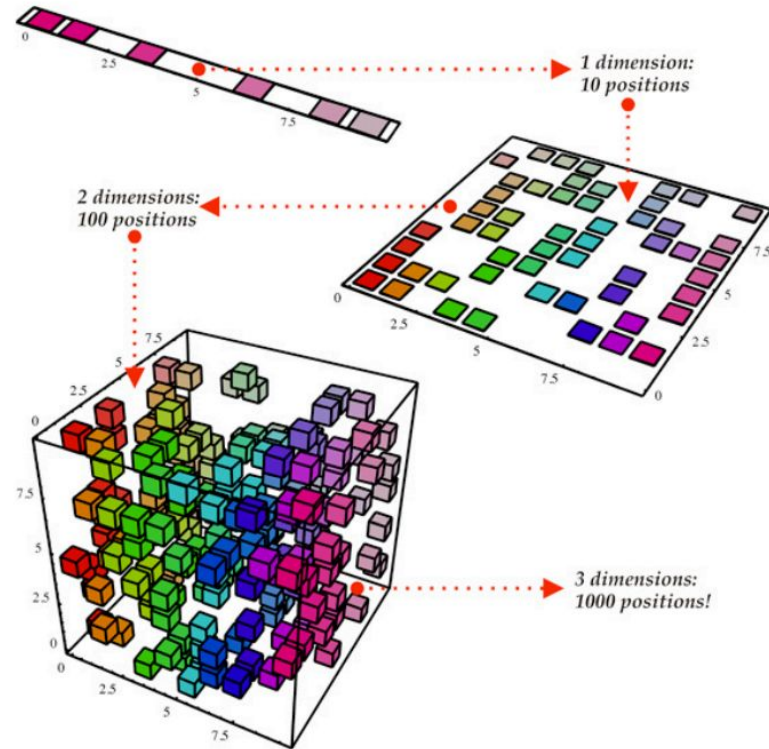


# The Concept of Unsupervised Learning - Clustering Tasks

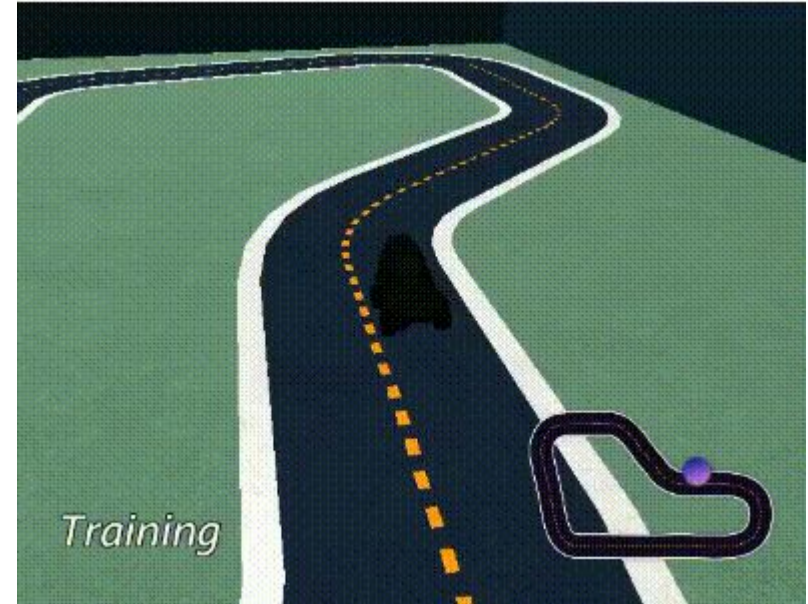
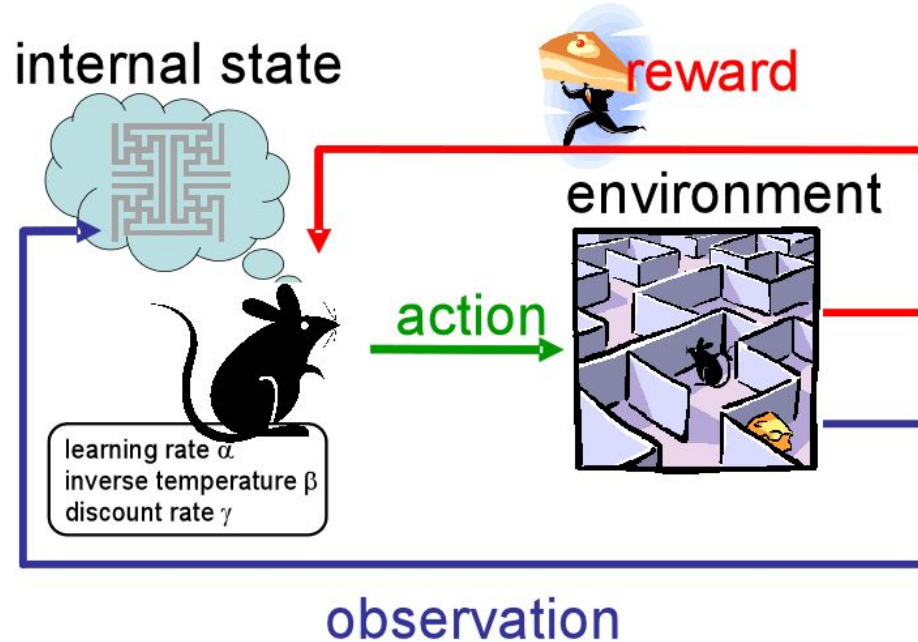




# The Concept of Unsupervised Learning - Dimensional Reduction Tasks



# The Concept of Reinforcement Learning

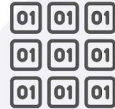




# Data Types in Machine Learning Context

# Structured vs. Unstructured vs. Semi-structured

## Structured data



### Characteristics

- Predefined data models
- Easy to search
- Text-based
- Shows what's happening

### Resides in

- Relational databases
- Data warehouses

### Stored in

- Rows and columns

### Examples

- Dates, phone numbers, social security numbers, customer names, transaction info

## Unstructured data



### Characteristics

- No predefined data models
- Difficult to search
- Text, pdf, images, video
- Shows the why

### Resides in

- Applications
- Data warehouses and lakes

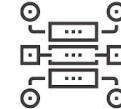
### Stored in

- Various forms

### Examples

- Documents, emails and messages, conversation transcripts, image files, open-ended survey answers

## Semi-structured data



### Characteristics

- Loosely organized
- Meta-level structure that can contain unstructured data

- HTML, XML, JSON

### Resides in

- Relational databases
- Tagged-text format

### Stored in

- Abstracts & figures

### Examples

- Server logs, tweets organized by hashtags, emails sorting by folders (inbox; sent; draft)

LEVITY



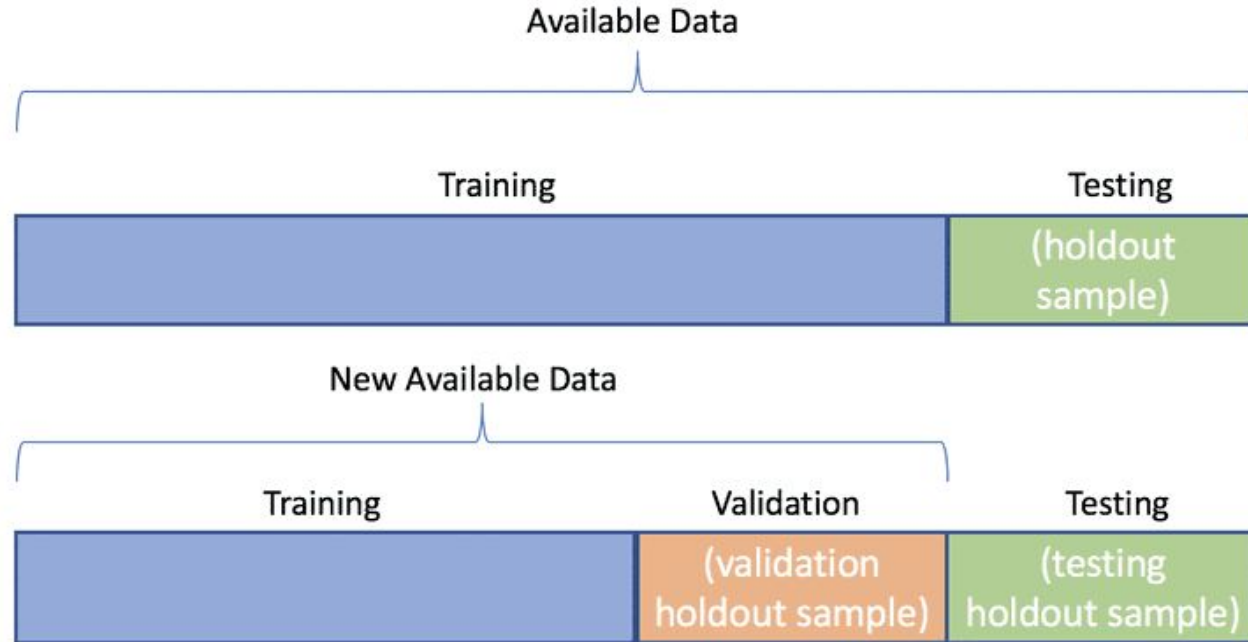
# Training Data, Testing Data, Validation Data



# Training vs. Testing vs. Validation Data #1

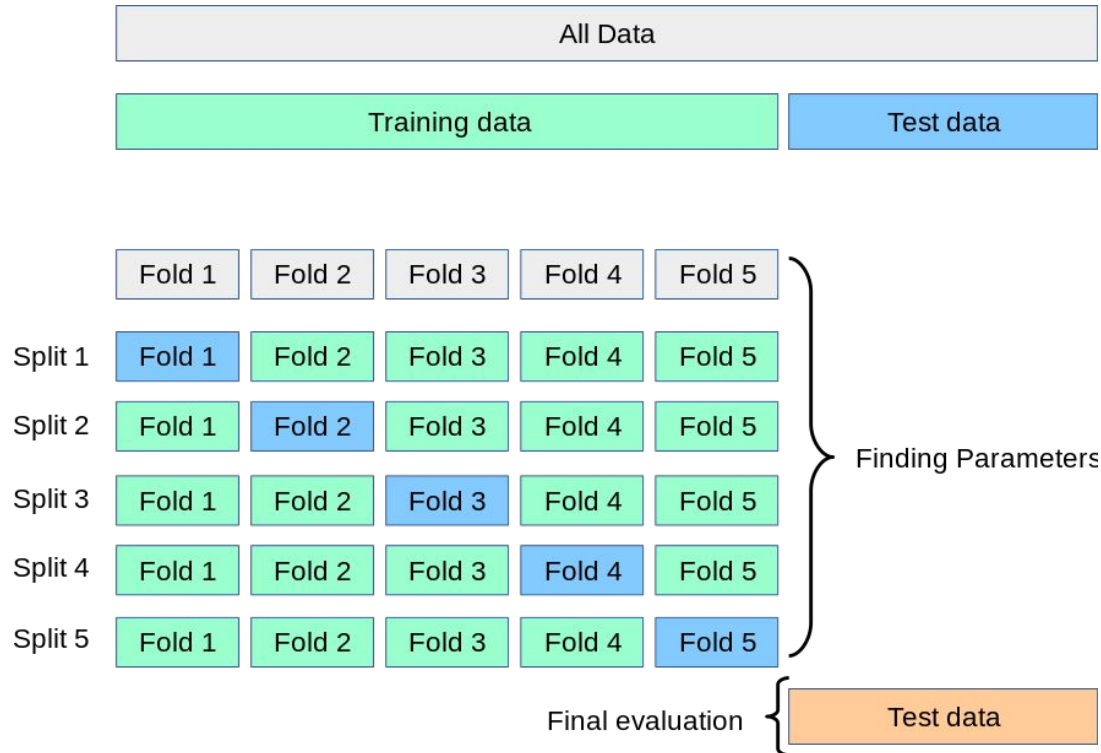
- It is common to partition a single set of supervised observations into training, validation, and test sets.
- No requirements for the sizes of the partitions.
- Vary according to the amount of data available.
- It is common to allocate between fifty and seventy-five percent of the data to the training set.
- Ten to twenty-five percent of the data to the test set.
- The remainder to the validation set.
- The validation set is used to tune variables called hyperparameters that control how the algorithm learns from the training data.

# Training vs. Testing vs. Validation Data #2





# Cross-validation Approach #1





# Cross-validation Approach #2





# Introduction to Machine Learning Model Evaluation

# How to Measure ML Model Performance? #1

- Performance metrics measure the amount of prediction error.
- **Bias**
  - An error from erroneous assumptions in the learning algorithm.
- **Variance**
  - A type of error that occurs due to a model's sensitivity to small fluctuations in the training set.

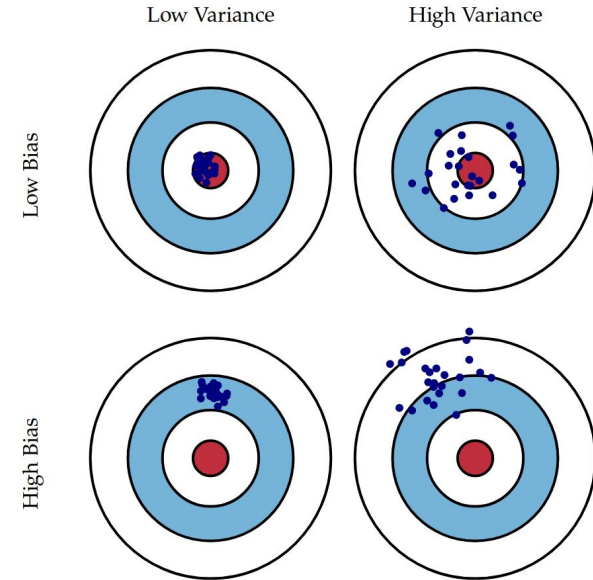
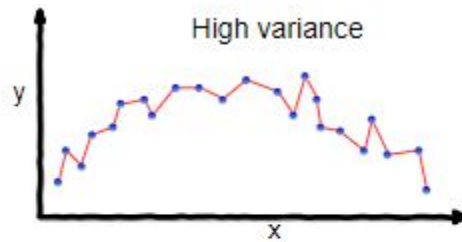
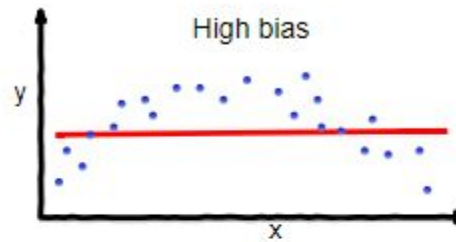


Fig. 1 Graphical illustration of bias and variance.

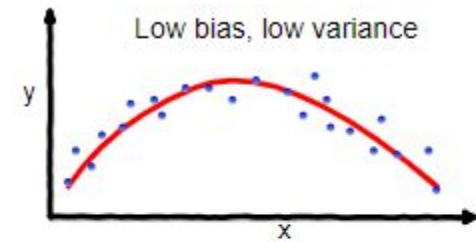
# How to Measure ML Model Performance? #2



overfitting



underfitting



Good balance

*The fundamental goal of machine learning is to generalize, or to induce an unknown rule from examples of the rule's application.*  
– Hackeling, 2017



# Setup Machine Learning Development Environment



# The Tools – Preferred in Our Class

- Code Editor – Visual Studio Code (later on, we will use Google Collab)
- Python
- Numpy
- Pandas
- Scikit-learn
- Pillow / OpenCV
- Tensorflow
- Joblib
- pyh5py
- Flask API

# Let's get your hands dirty!

## *Setup Environment Demonstration*