



Data Driven Research Using Machine Learning

Professor Swakkhar Shatabda
Dept of Computer Science and Engineering
BRAC University

Data, Information and Knowledge

32

Data, **Information** and Knowledge



32 °C | °F

Data, Information and **Knowledge**



32

°C | °F

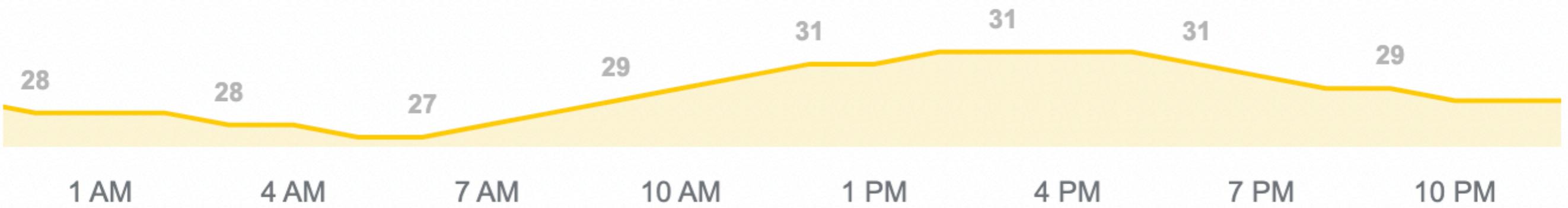
Precipitation: 60%
Humidity: 72%
Wind: 14 km/h

Dhaka

Wednesday

Scattered thunderstorms

Temperature | Precipitation | Wind



Tue



32° 27°

Wed



32° 27°

Thu



32° 28°

Fri



32° 28°

Sat



34° 28°

Sun



34° 28°

Mon



33° 28°

Tue



33° 28°

Data

Raw



Information

Processed



Knowledge

Actionable

Authoritative Capacity



Past Experience

Decision

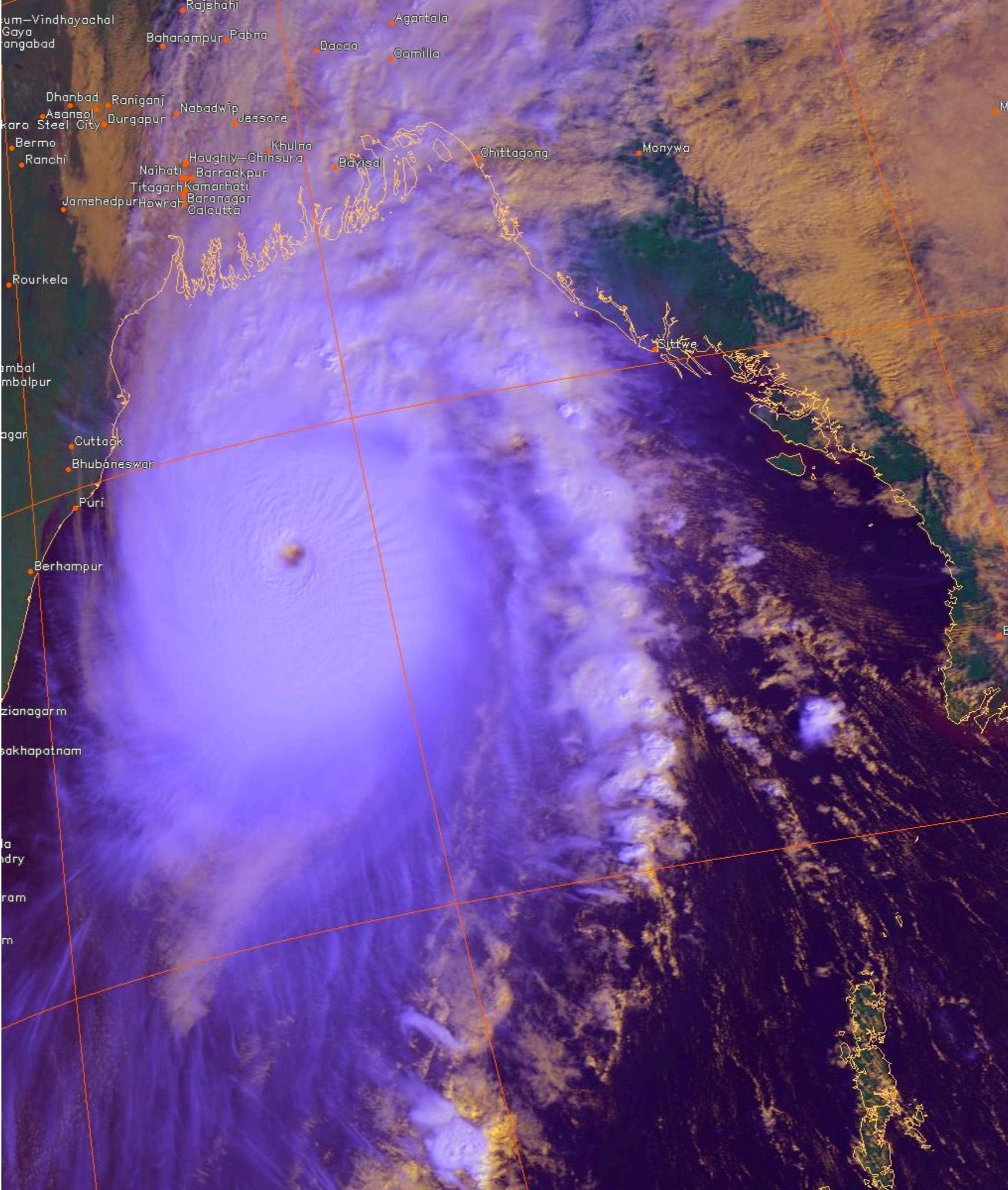
Making

Adaptive

Unstructured Data

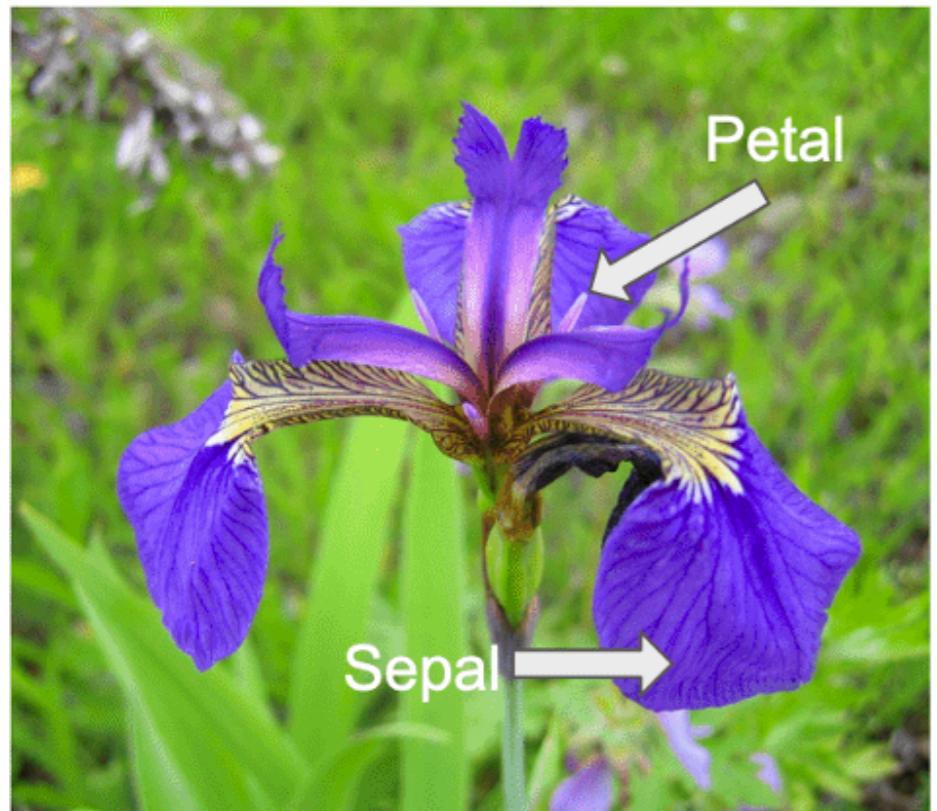


Unstructured Data

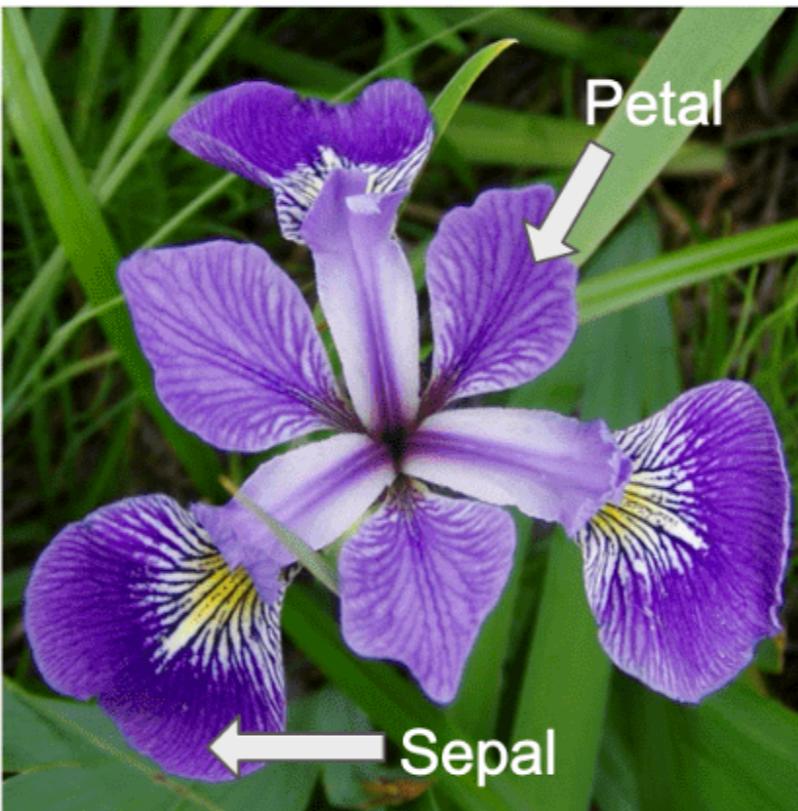


Iris Classification

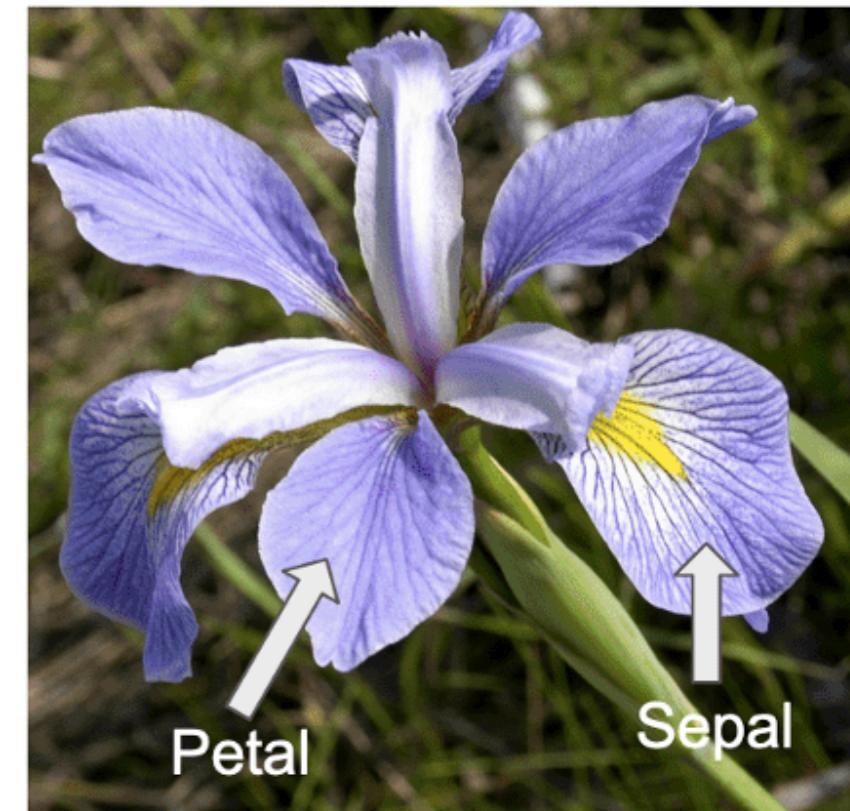
Iris setosa



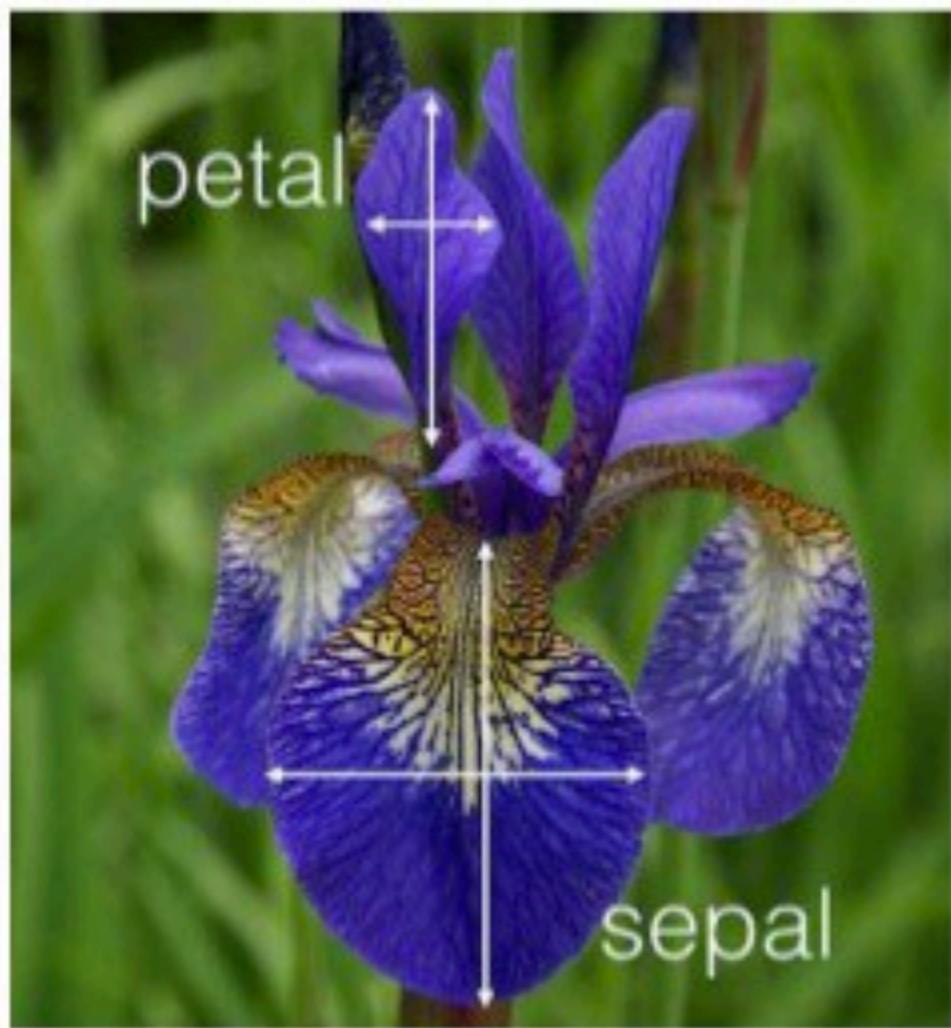
Iris versicolor



Iris virginica



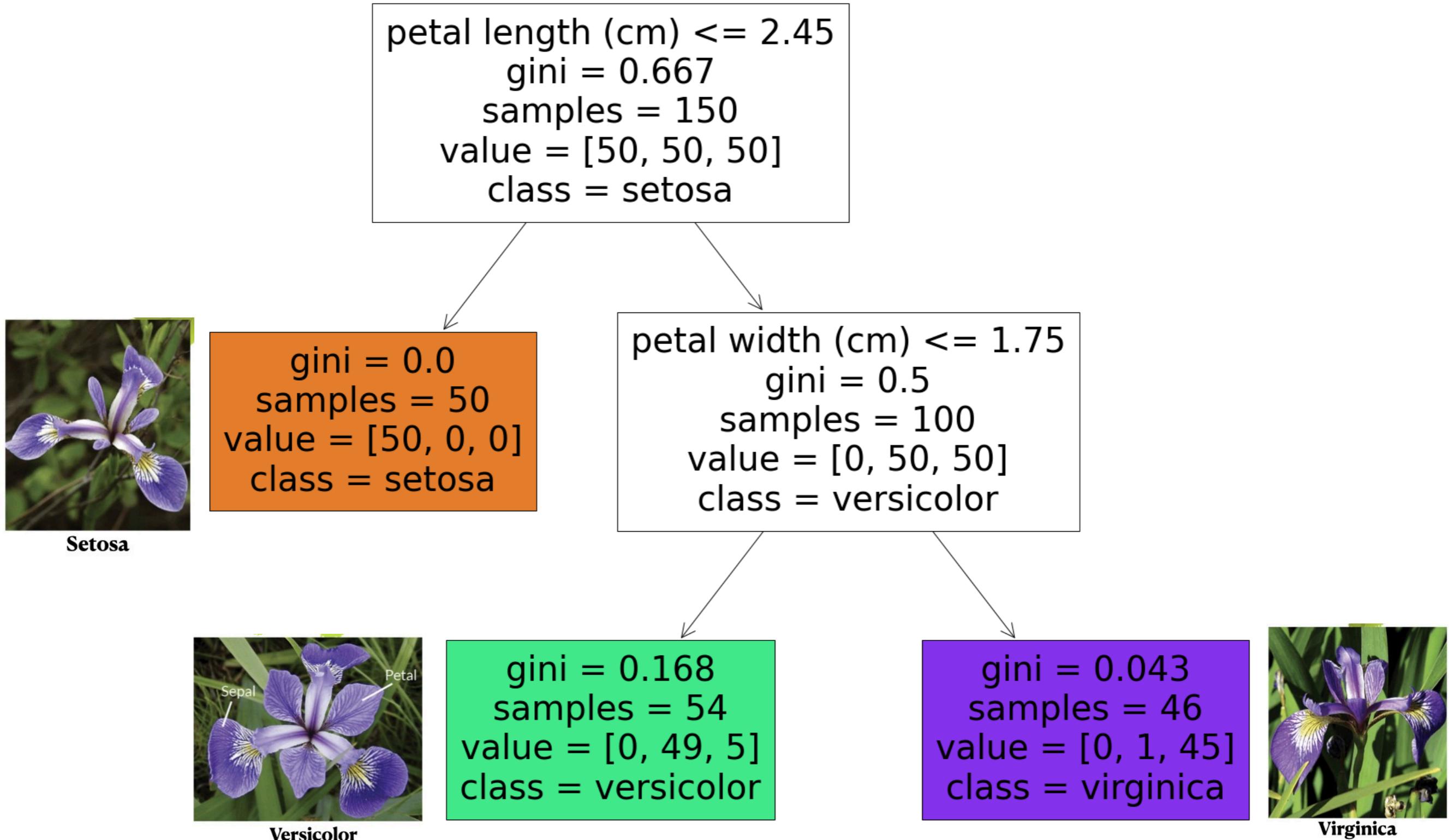
Structured Data



	sepallength	sepalwidth	petallength	petalwidth	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

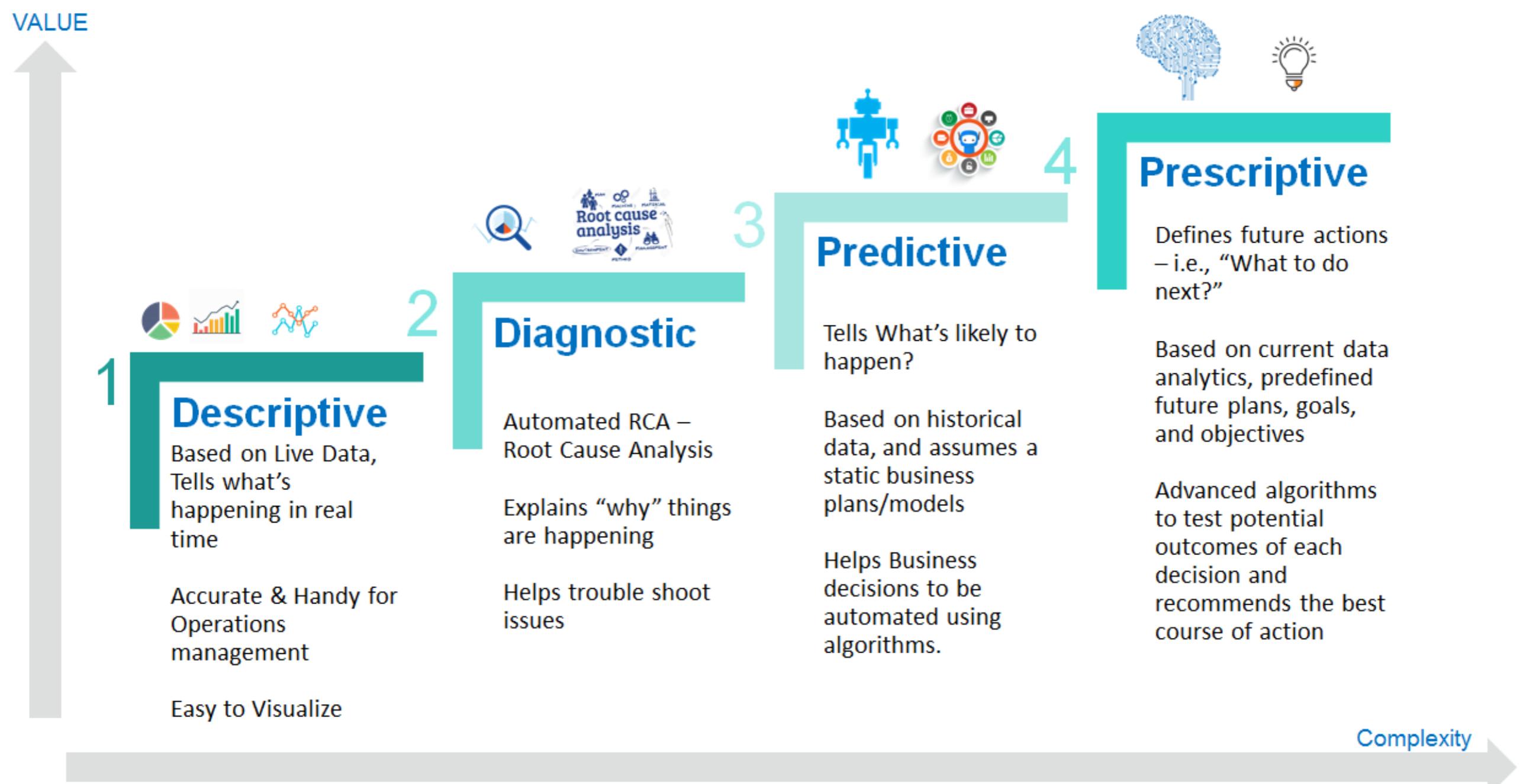
150 rows × 5 columns

Structured Data > Knowledge



Data Analytics!

Analytics is the systematic computational analysis of data or statistics. It is used for the discovery, interpretation, and communication of meaningful patterns in data. It also entails applying data patterns toward effective decision-making.



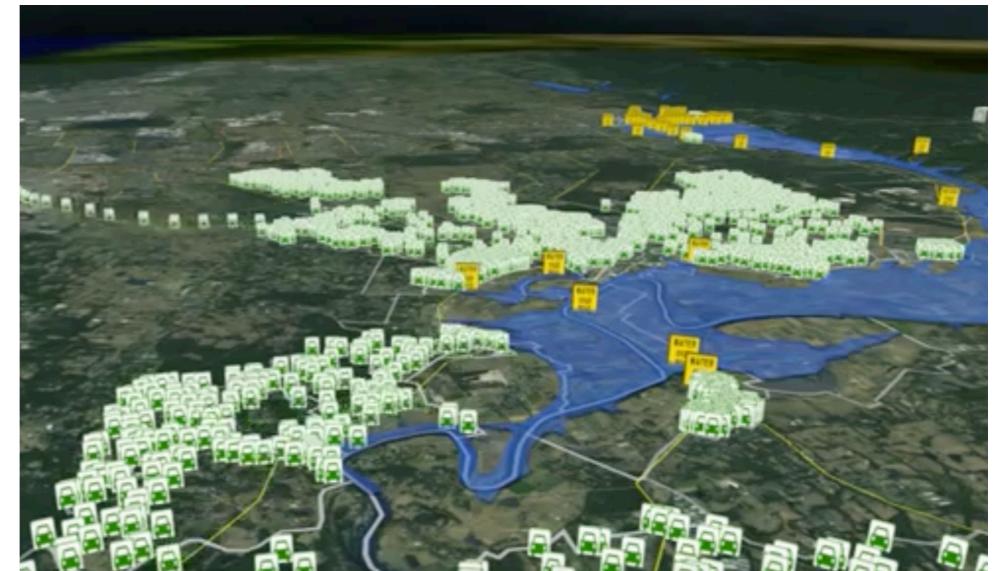


Problem First:
Relocating People

Then look for data

Relocating People: **Problem First**

- Service Innovation
 - When is the proper time to start relocating a household?
 - At what hours? Time
 - Which way to move a household is the best?
 - Highway 1 / Highway 2 / ...



Relocating People: Need Data

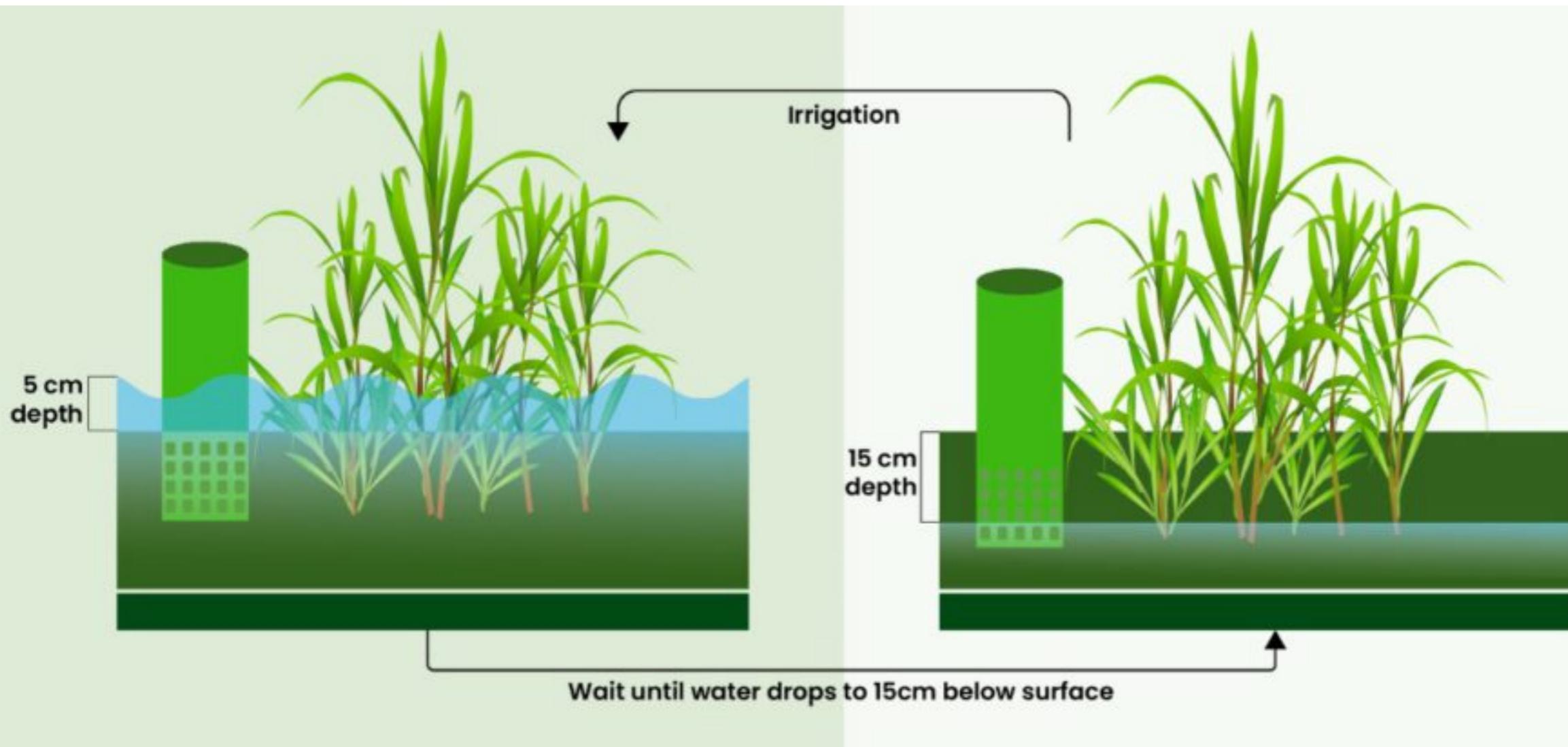
- Household information
 - Demographics
 - Location
 - Distance from river / highway / shelter
- Area Map
 - Road Distances
 - Weather Information



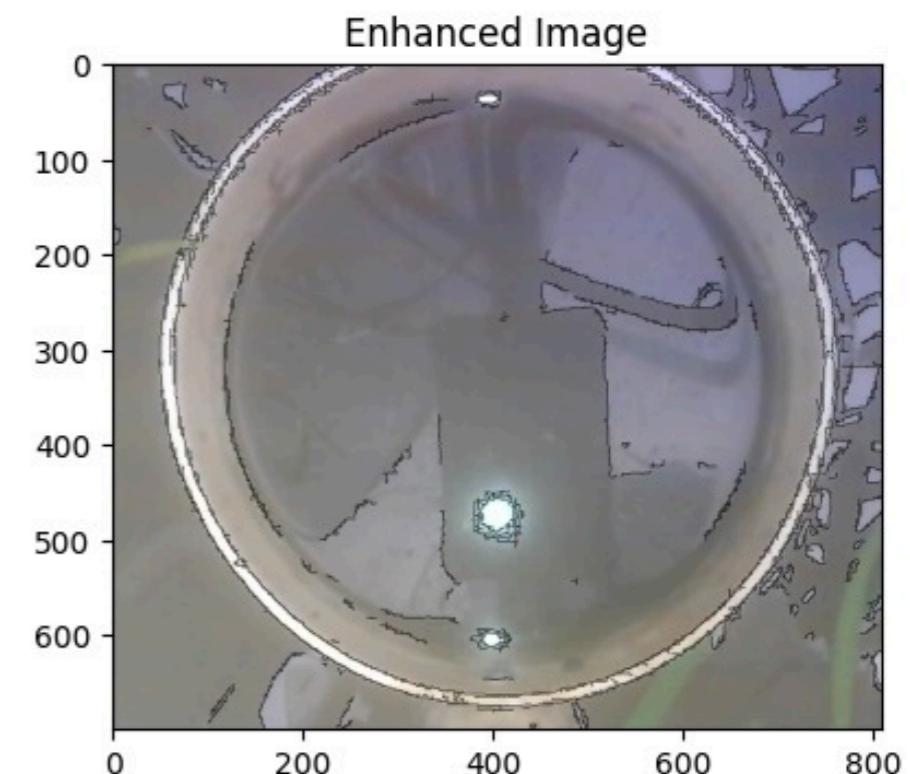
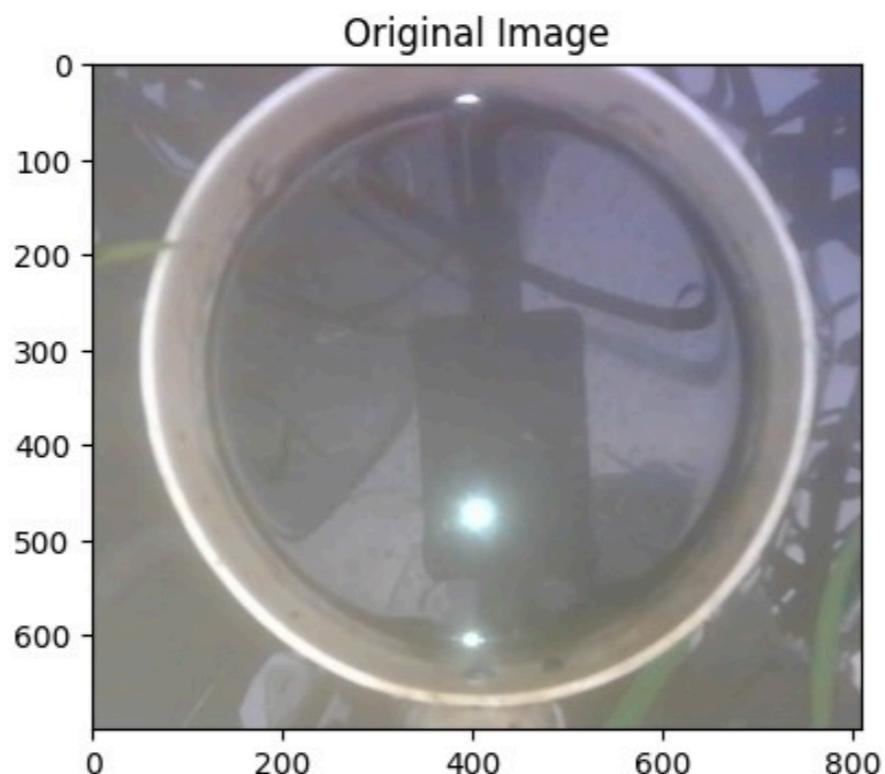
Water and Irrigation



Water and Irrigation



Water and Irrigation

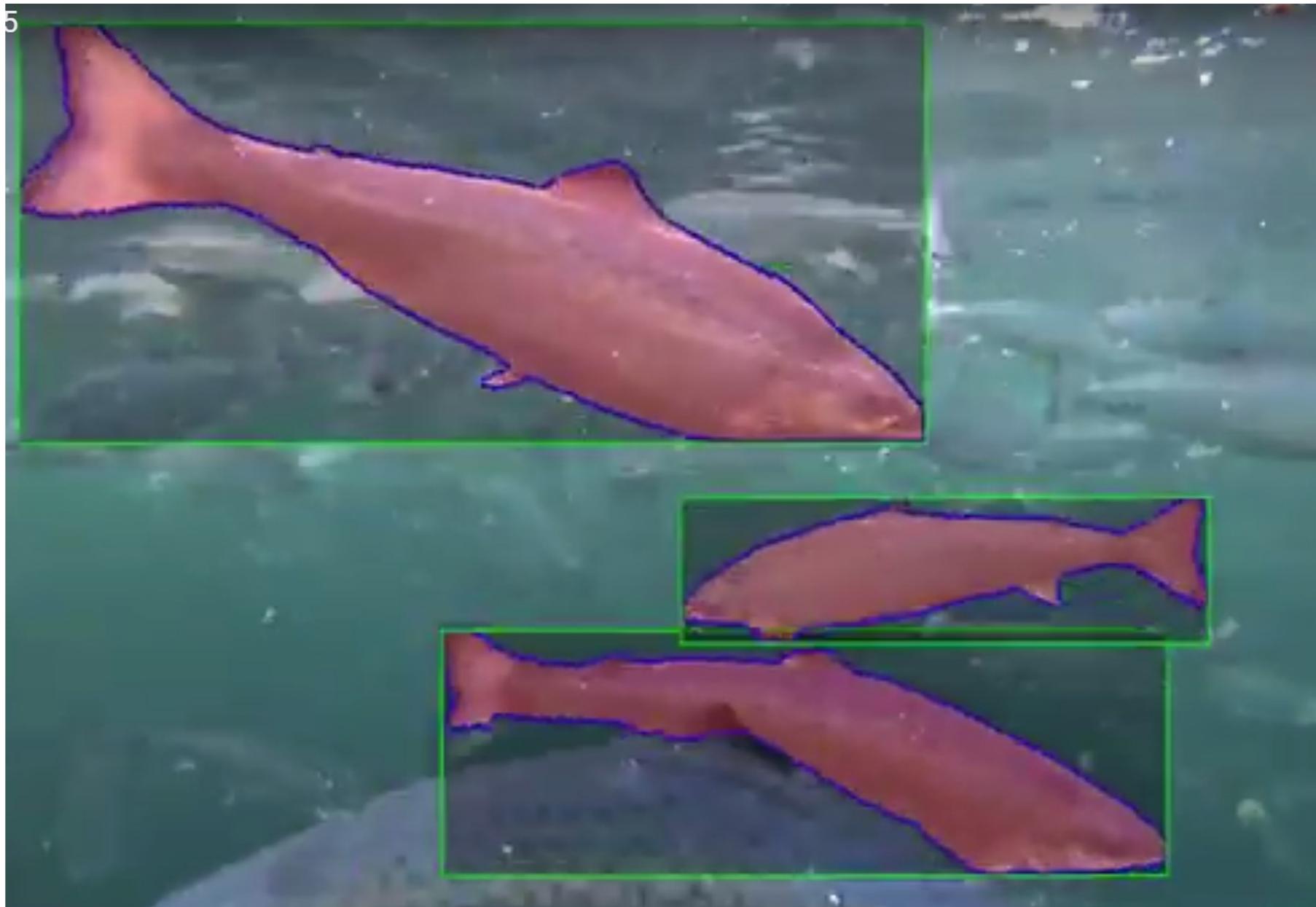


Salmon Farming



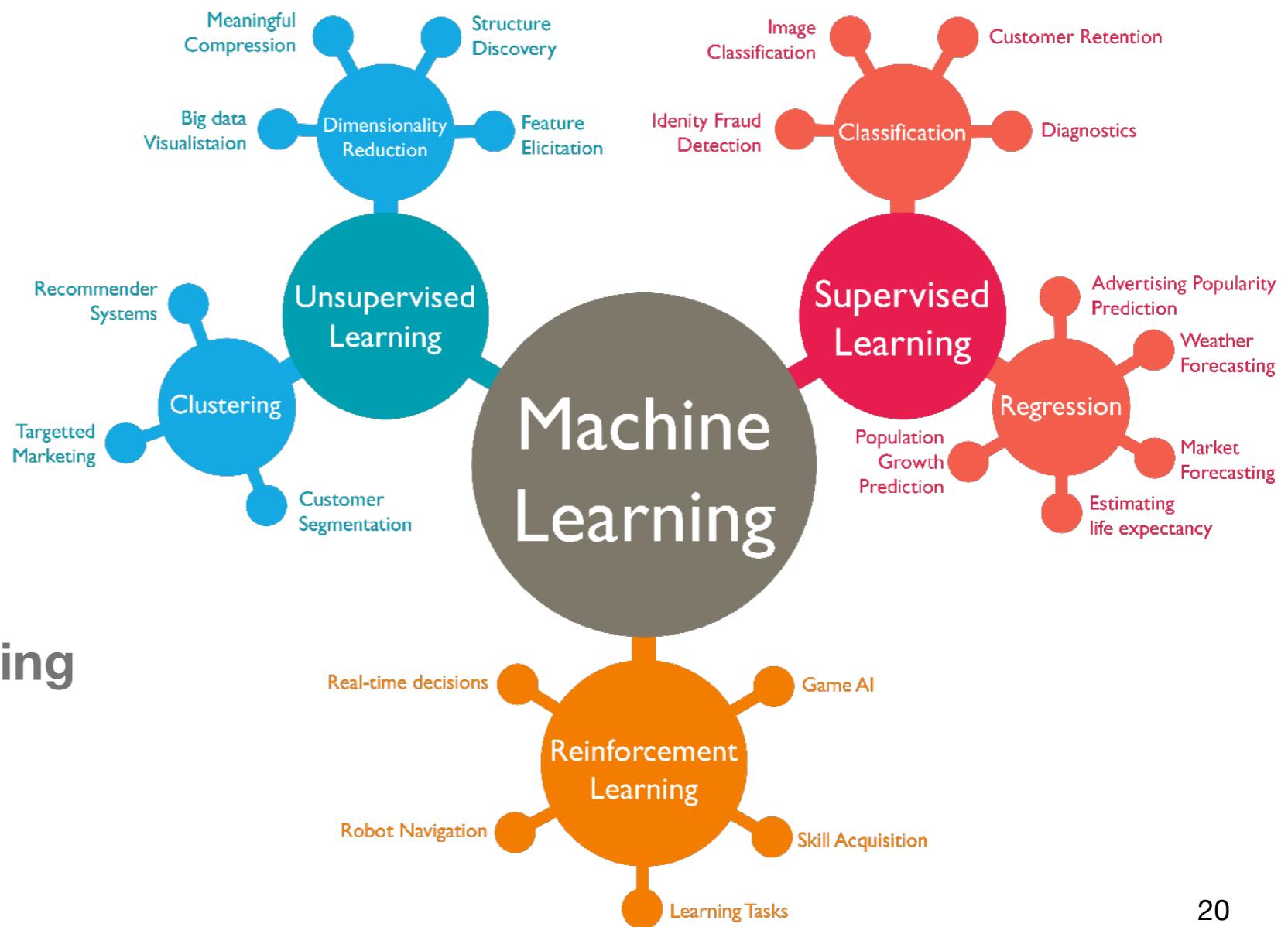
Salmon Farming

- Data Collection
- Segmentation
- Pre-processing
- Model Training
- Implementation



Machine Learning

- A computer program is said to learn from experience **E** with respect to some class of task **T** and performance measure **P**, if its performance at task T as measured by P improves with experience E.
 - Supervised Learning
 - **Classification**
 - Regression
 - Unsupervised Learning
 - **Clustering**
 - **Semi-Supervised Learning**
 - Reinforcement Learning

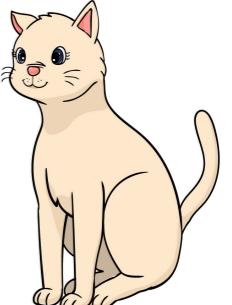


Classification

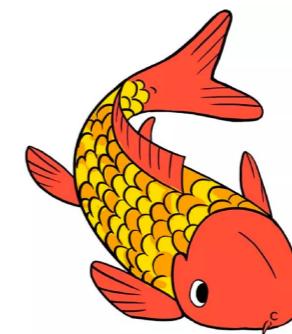
- Classification **maps data into predefined groups** or classes.
- It is often referred to as **supervised learning** because the classes are determined before examining the data.
- Classification creates a function from training data. The training data consist of **pairs of input objects, and desired output**.
- The output of the function can be a continuous value, or can predict a class label of the input object.



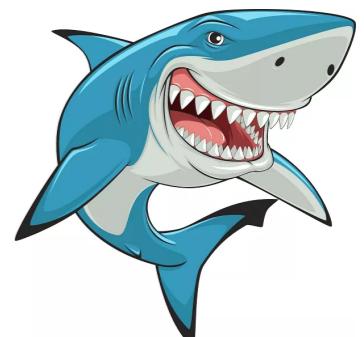
class A



class B



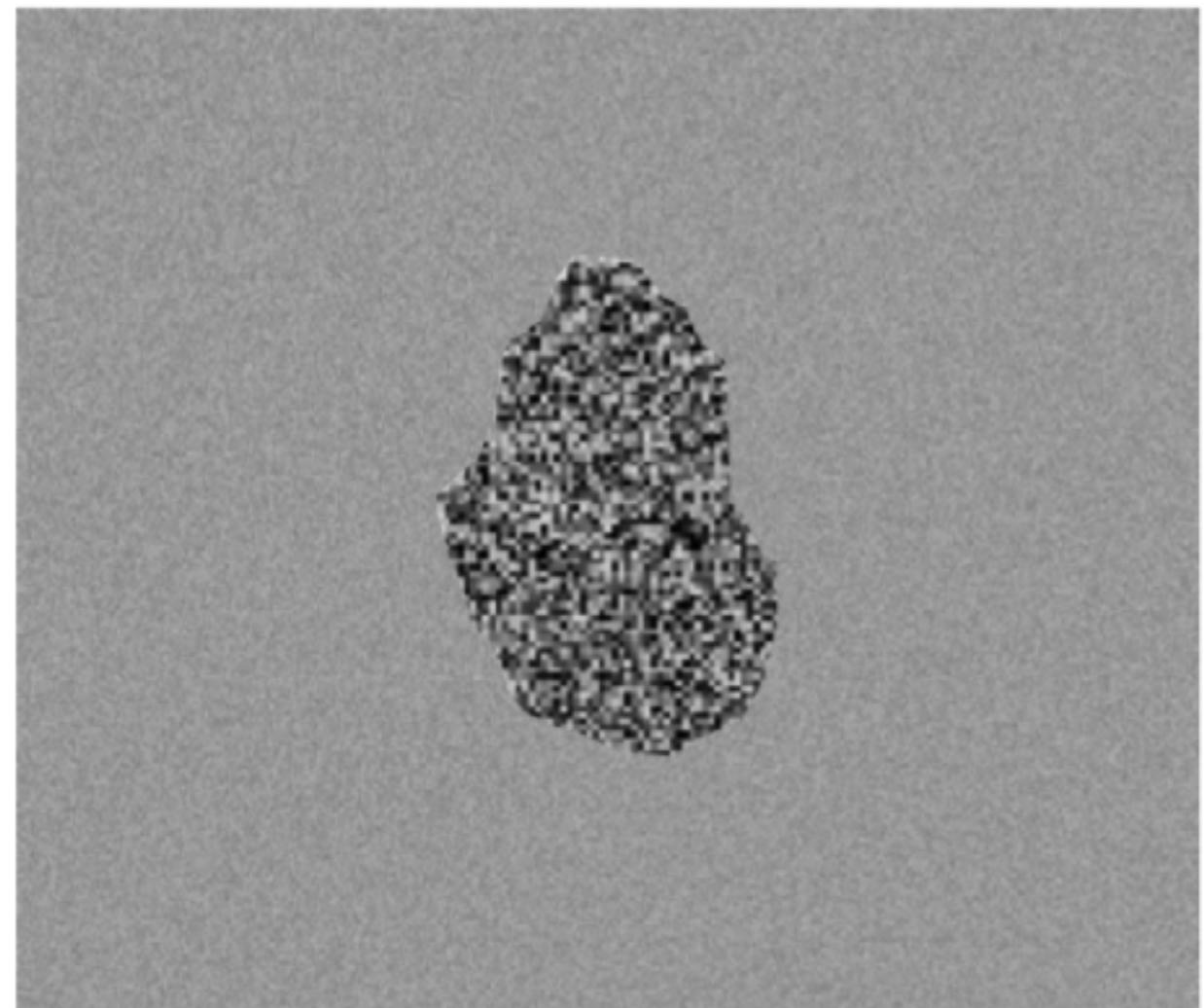
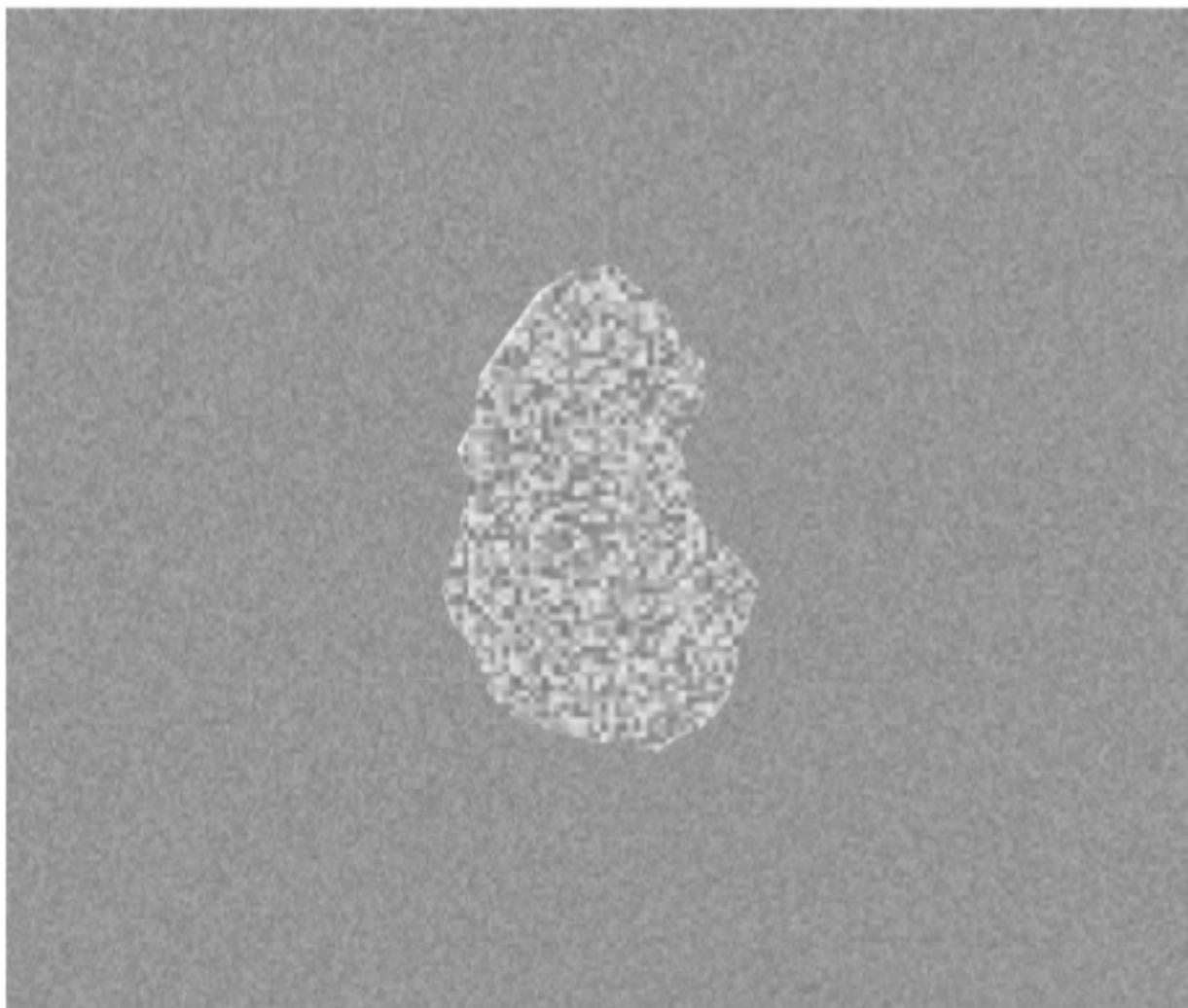
class C



class D

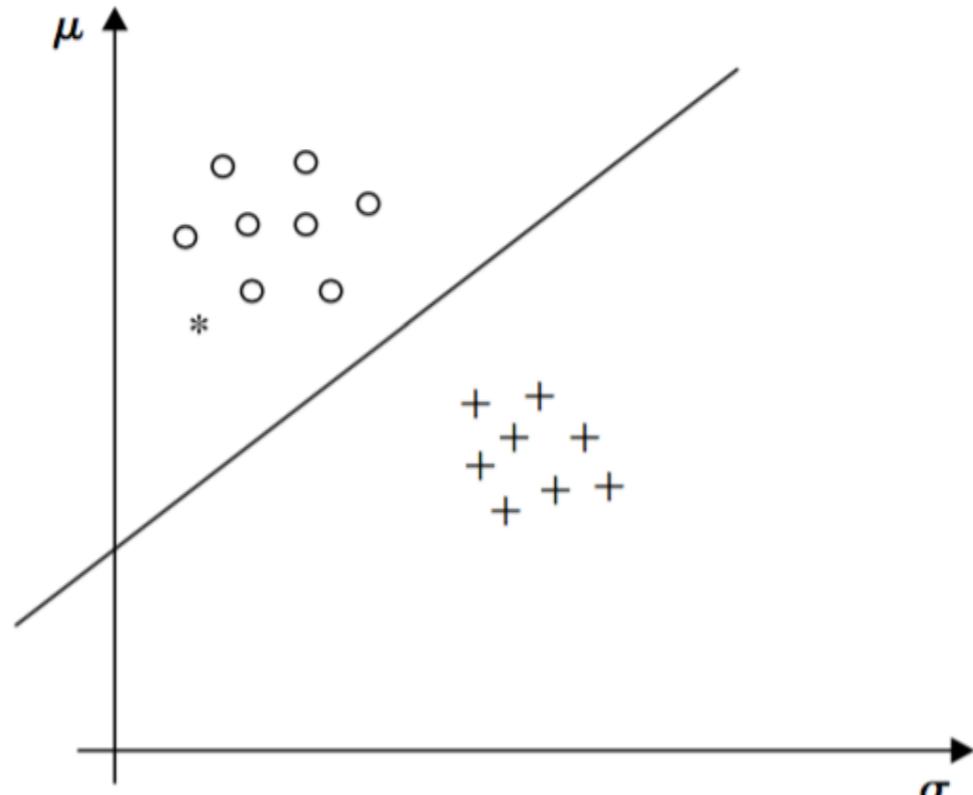
Medical Image Analysis

- Two classes: benign and malignant
- Identify the measurable quantities that make these two regions distinct from each other.



Features

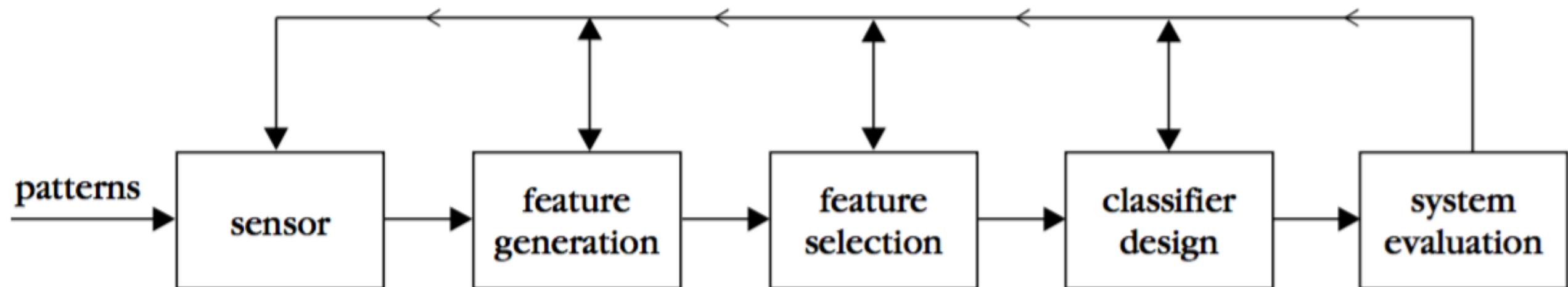
- **Intensity** in each region of interest versus the corresponding standard deviation around this mean.
- **Features:** the measurements used for the classification, a random variable (attributes/properties)
- Features are treated as vectors
- Each of the feature vectors identifies uniquely a single pattern (object/instance).
- **Decision line** constitutes the classifier



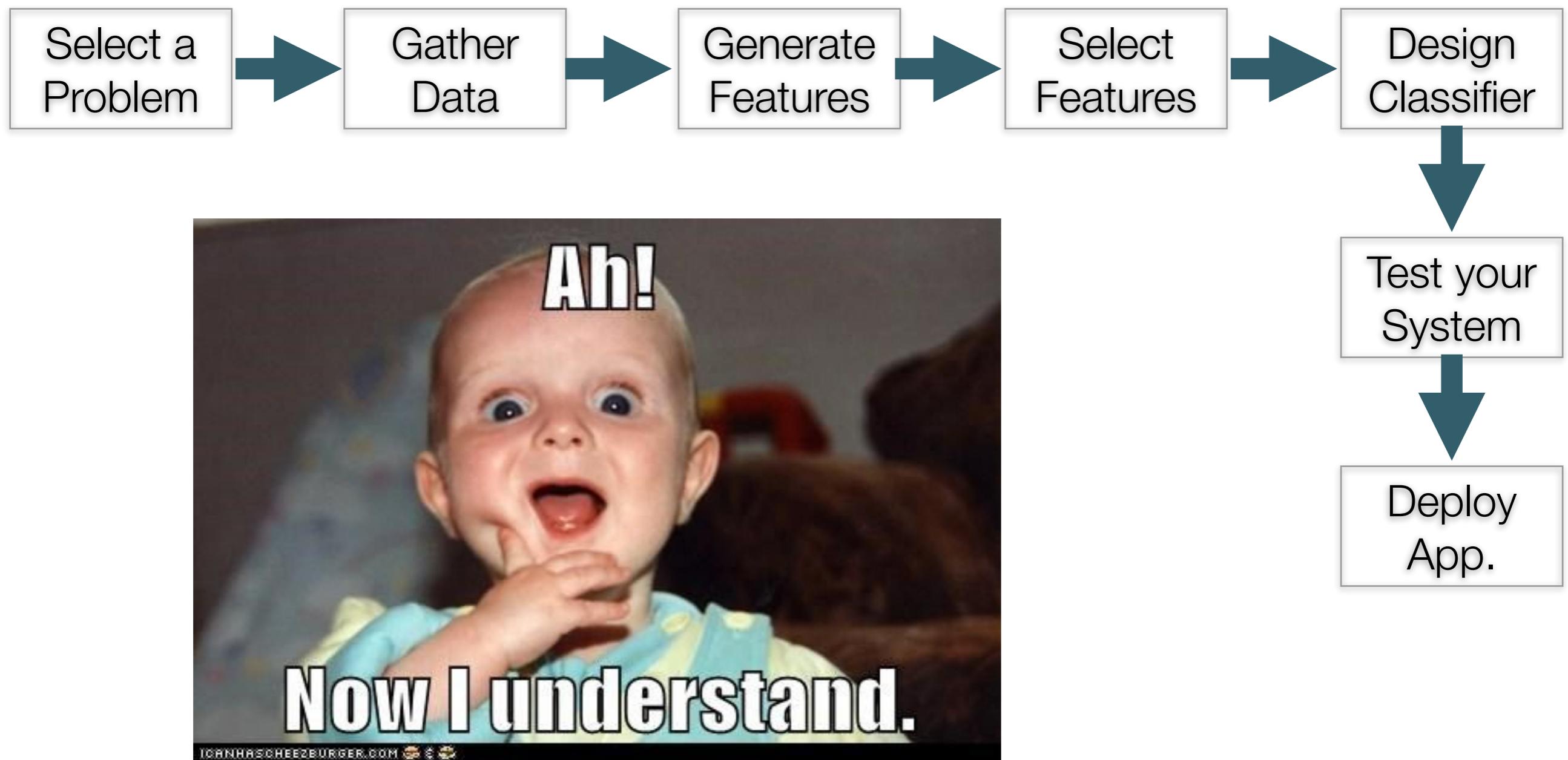
Plot of the mean value versus the standard deviation for a number of different images originating from two classes. In this case, a straight line separates the two classes.

Classifier Design

- **Feature Generation:** How are the features generated?
- **Feature Selection:** What is the best number l of features to use?
- **Classifier Design:** How does one design the classifier?
- **System Evaluation:** How can one assess the performance of the designed classifier?



Classification System



Problem Formulation

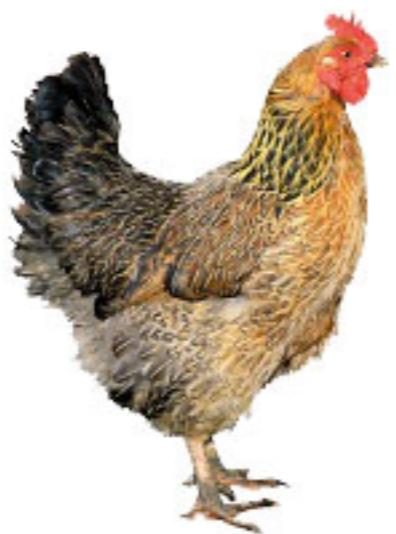
- To find the correct chicken!



DESHI VS BIDESHI

Gather Data?

DESHI



BIDESHI



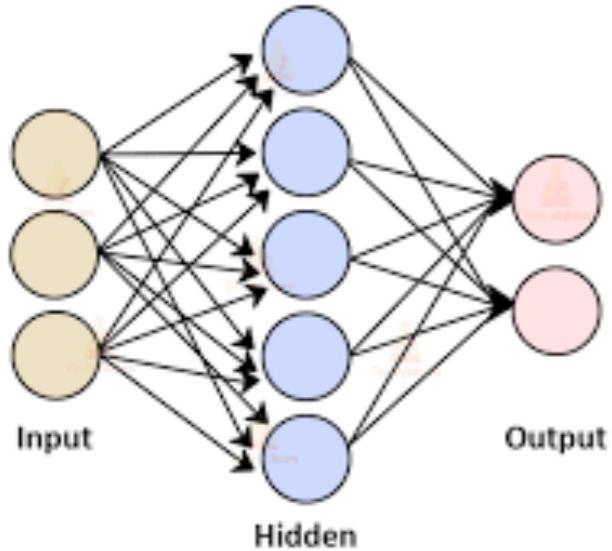
- We can use a smart phone / sensor /

Featurization: Generation and Selection

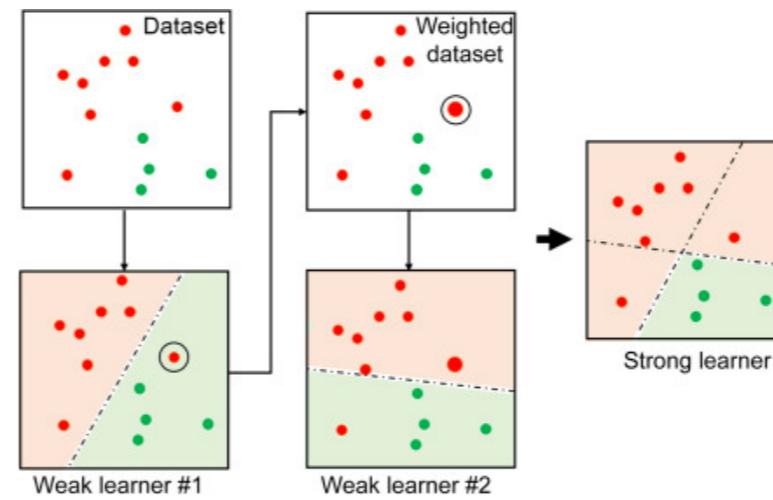
- Color
- Comb
- Hackle
- Saddle
- Claw
- Sickle Feather
-
- And if nothing pixels!



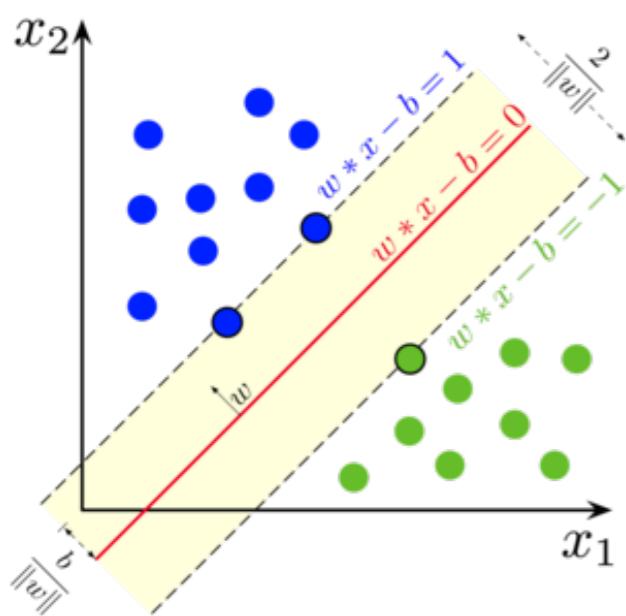
Classifier Design



Artificial Neural Networks



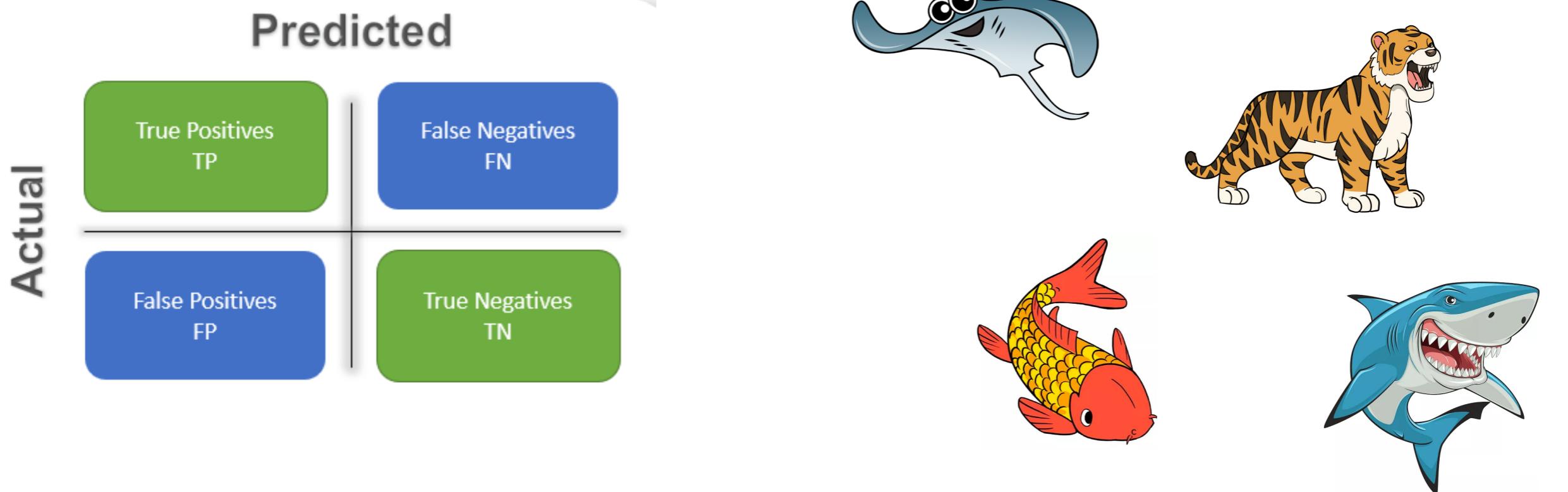
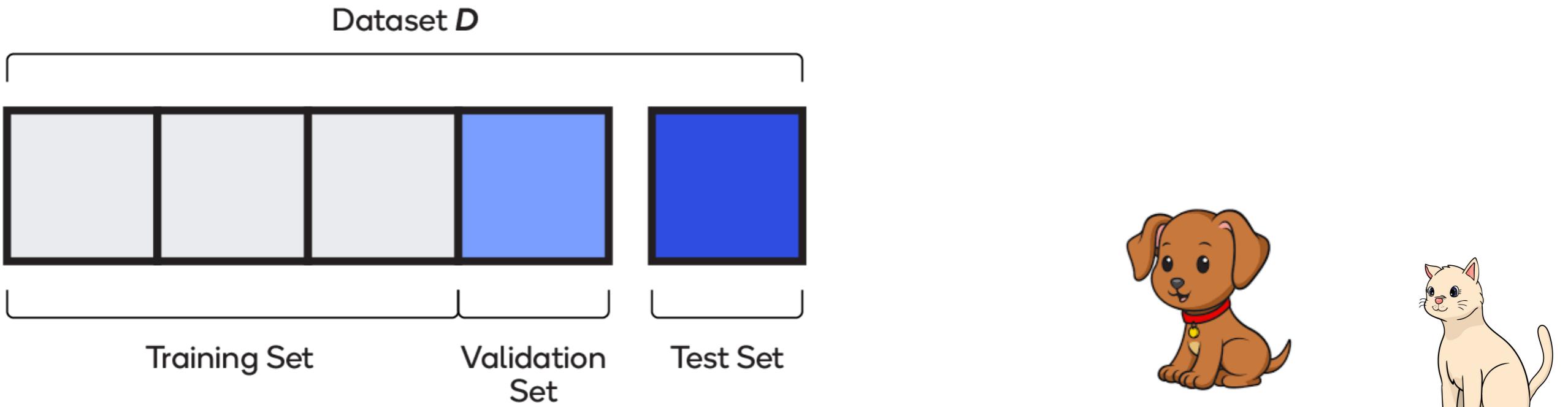
AdaBoost Classifier



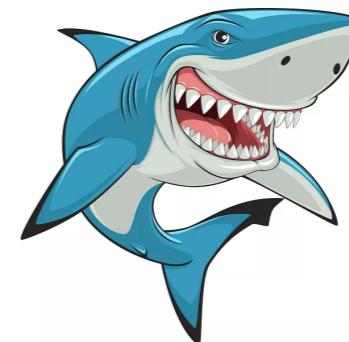
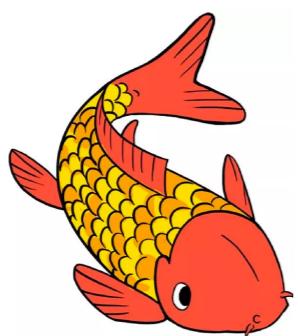
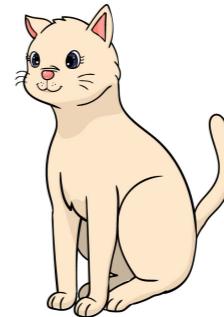
Support Vector Machine



Evaluation



Fish Detection System



Yes

No

Deployment

- Make the app usable for others
- Intelligent System



Deep Learning

- Deep-learning methods are **representation-learning** methods with **multiple levels** of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level....The key aspect of deep learning is that these layers of **features are not designed by human engineers**: they are learned from data using a general-purpose learning procedure.



Y. Bengio

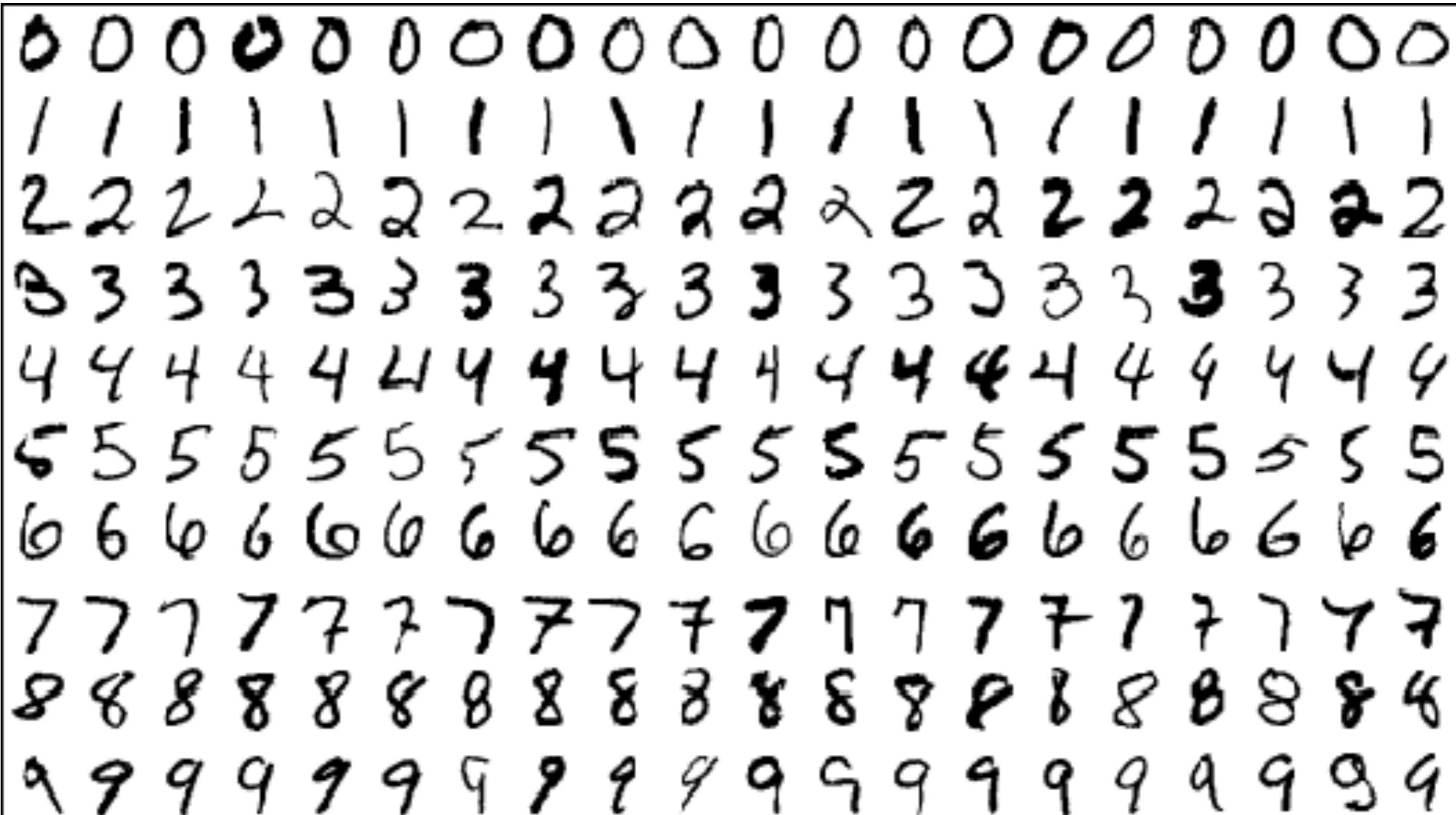


Geoffrey Hinton

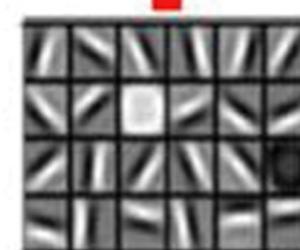
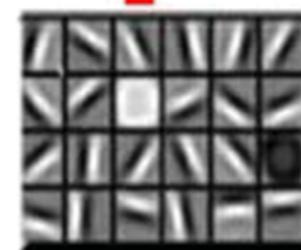
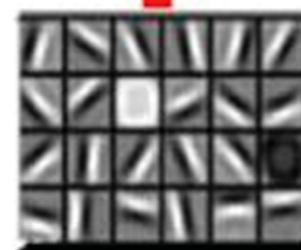
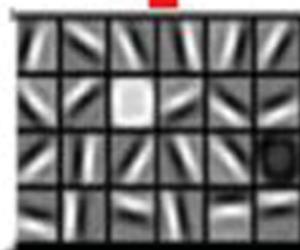
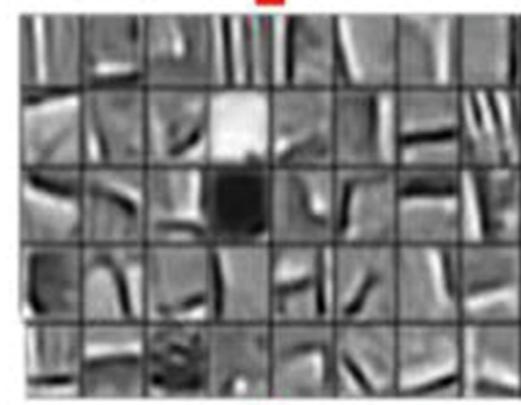
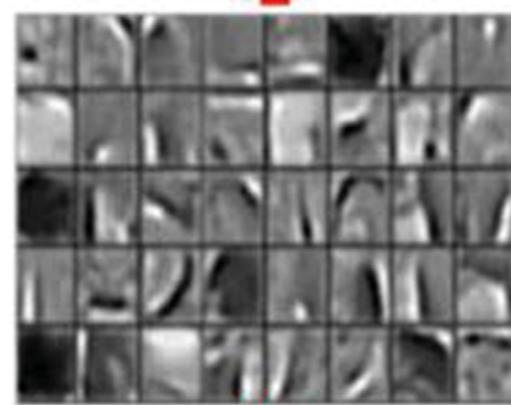
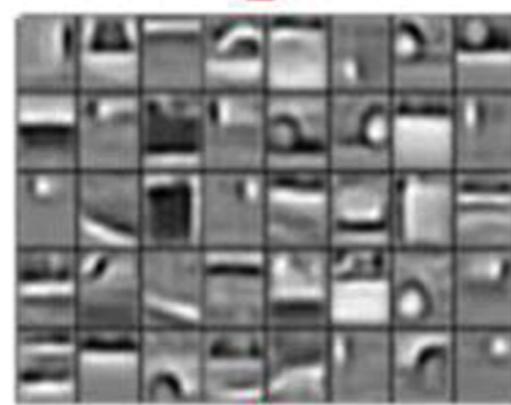
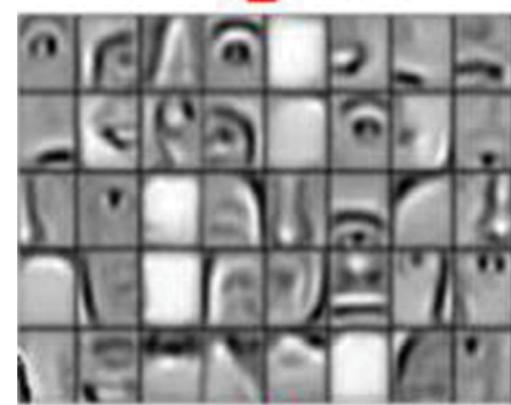
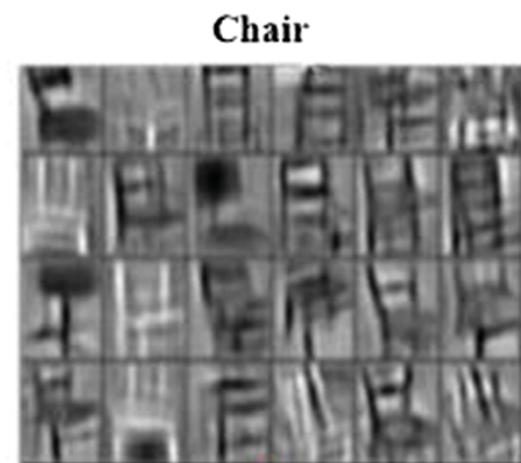
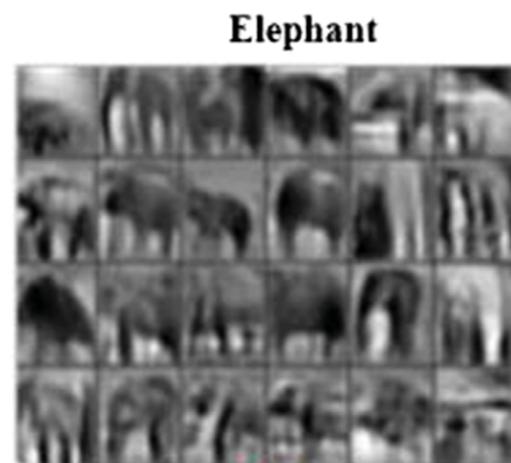
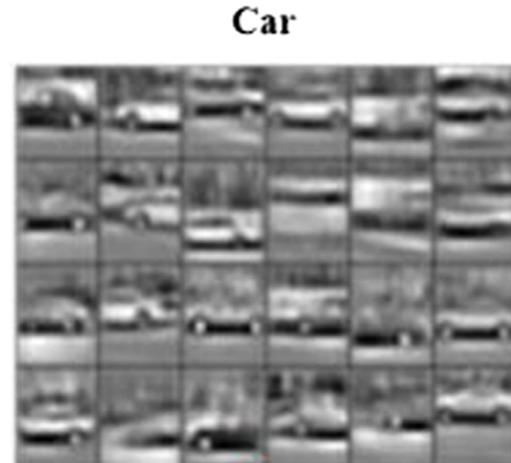
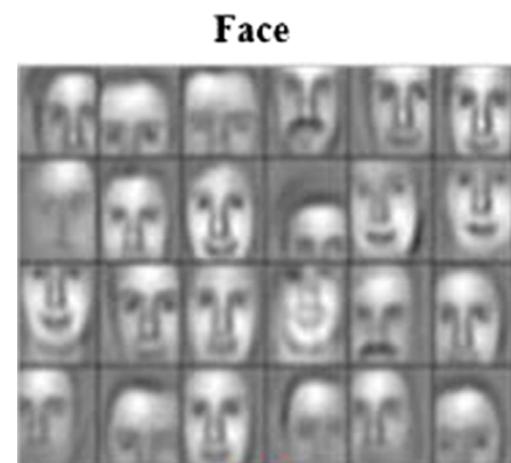


Yann LeCun

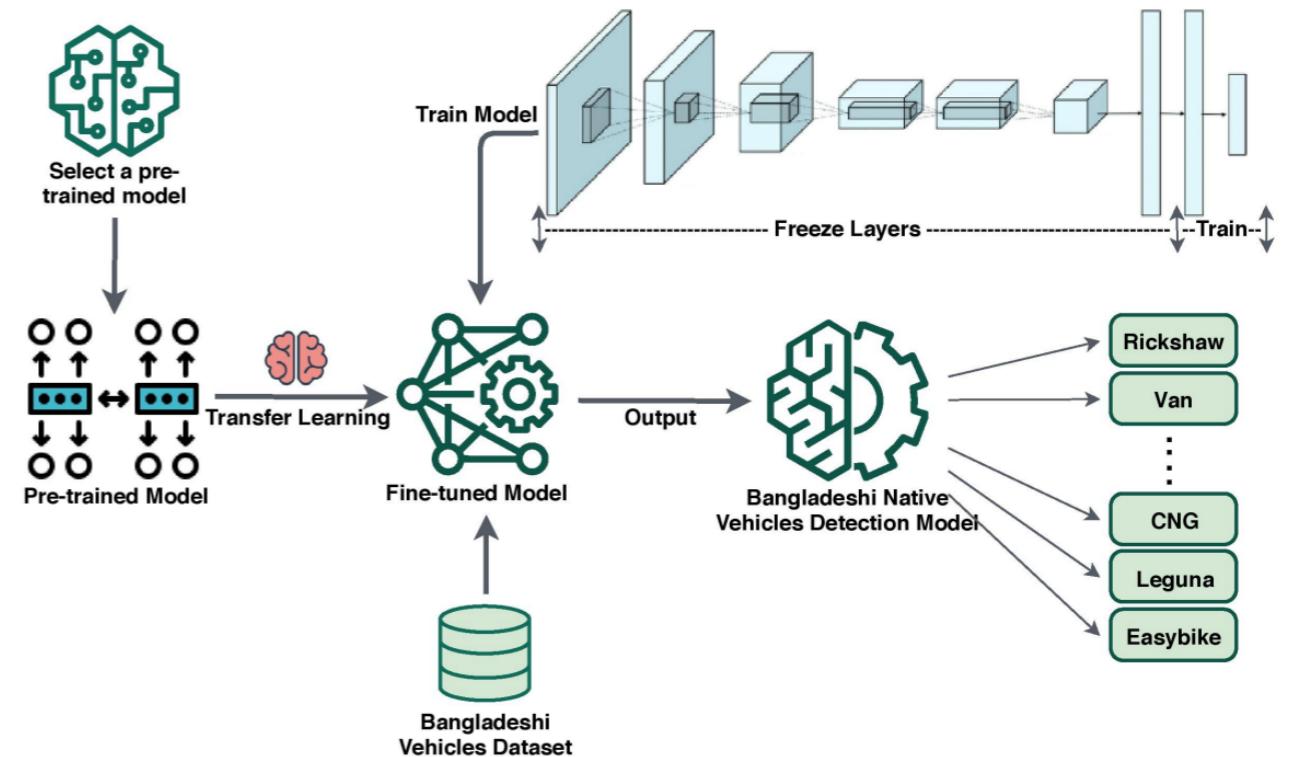
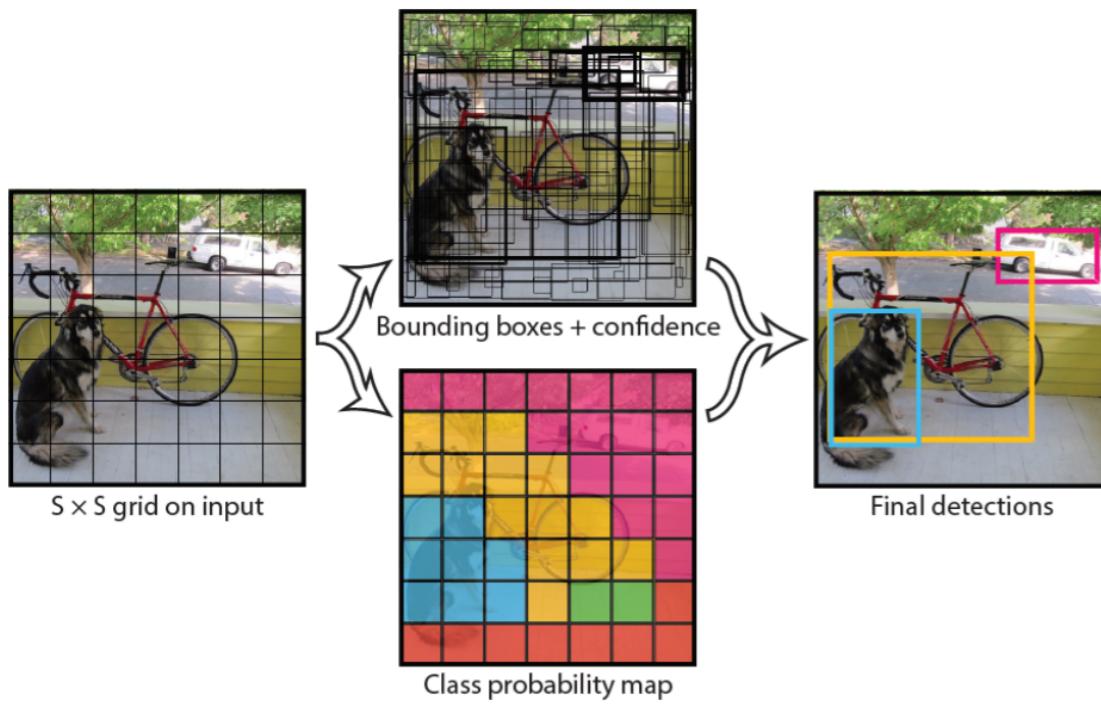
Deep Learning



Deep Feature Representation



Transfer Learning: Native Vehicle Detection



Existing Model: Yolo



Transfer Learning: Workplace Behaviour

