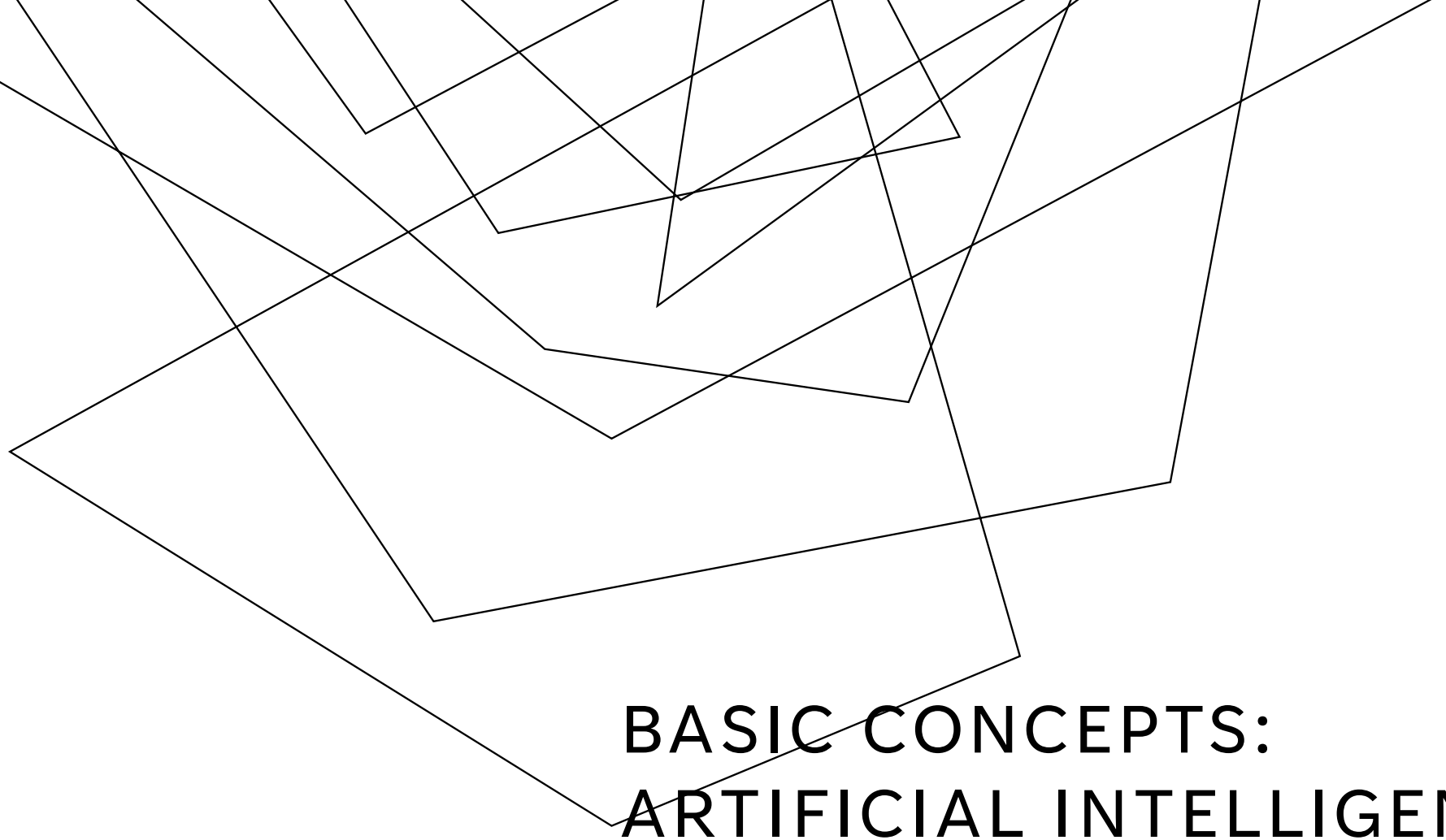


BASIC CONCEPTS: ARTIFICIAL INTELLIGENCE DEPLOYMENT

Eduardo Morales

MACHINE LEARNING ALGORITHMS

Supervised learning

- Object detection
- Image classification
- Weather estimation

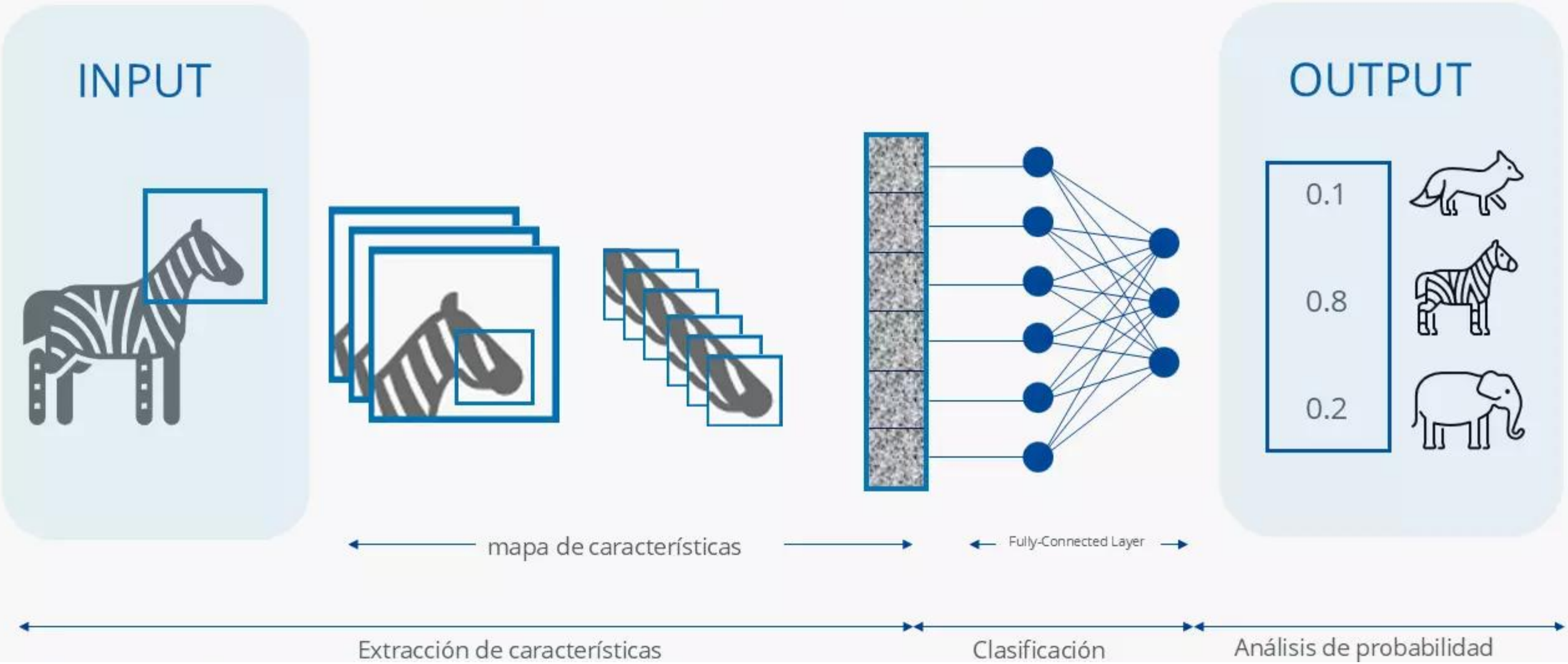
Non supervised learning

- Recommendation systems
- Fraud detection

Reinforcement learning

- Robotics
- Videogames
- Resource management

CONVOLUTIONAL NEURAL NETWORK (CNN)



OBJECT DETECTION

Bounding boxes + confidence

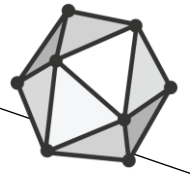
Class probability map

YOLO

Final detections

you only look ONCE

p_o	- Object probability
x	- x coordinate
y	- y coordinate
w	- width
h	- height
c_1	- one-hot class 1
c_2	- one-hot class 2
c_3	- one-hot class 3



ONNX

Hugging Face

ALL THOSE MODELS HAVE A LOT OF REQUIREMENTS TO TRAIN OR TO MAKE INFERENCE (PREDICT A RESULT)

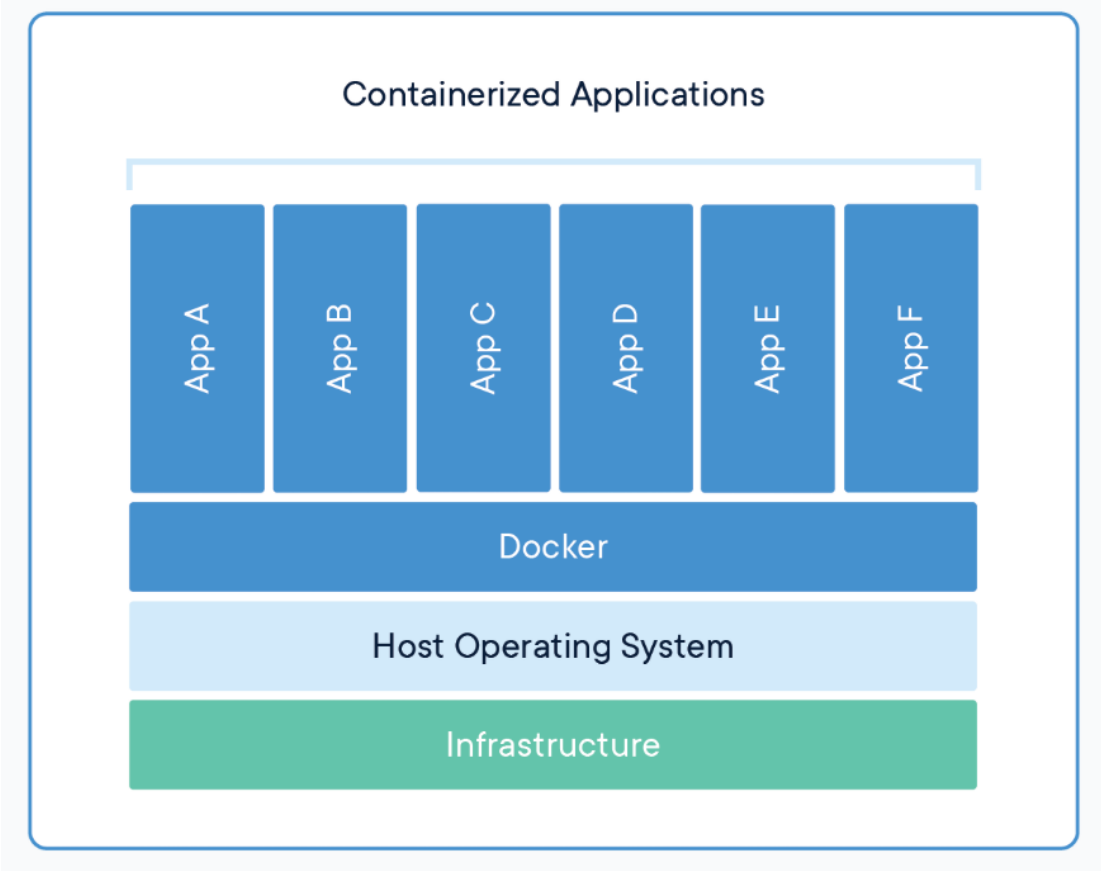
TensorFlow

Dopamine

DOCKER

- Allow developers to deploy artificial intelligence models in containers

- Self contained virtual machines
  - Libraries
  - Configurations
  - Frameworks

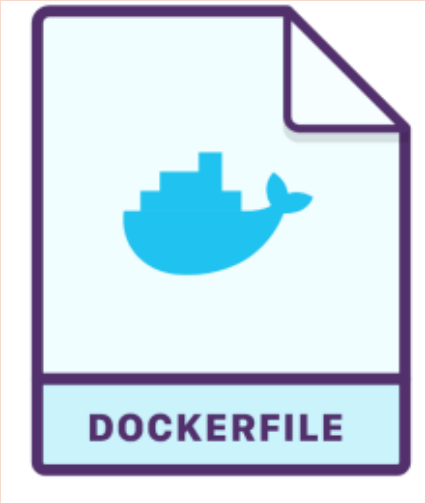


ENABLE

- Consistency
  - All the containers are executed without installing the requirements. Minimizing compatibility problems.
- Efficiency
  - Compared to virtual machines, the kernel is reused for all containers.
- Scalability
  - Enable replication and modification of environments.
- Portability
  - The application can be executed in nearly any environment.

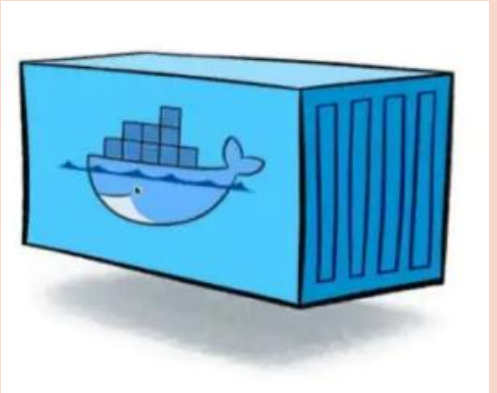
COMPONENTS

Docker file



Docker image

Docker container



ACTIVITY

- We are going to deploy machine learning models using pre-build containers.
  - Image classification
  - Language model (Ollama)

  - Step 1 - > Install Docker desktop (www.docker.com/products/docker-desktop)
  - Step 2 - > Enable virtualization in your BIOS
  - Step 3 - > Download the repo (https://github.com/emoralesv/TC2008B)
    - It contain the docker files and images to perform inference

ACTIVITY

- docker pull ollama/ollama
- docker pull emoralesv/ml
- LLM
  - docker run --name ollama --gpus all –p 11434:11434 ollama/ollama
  - docker exec –it ollama bash
  - ollama pull llama3
- Image classifier
  - docker run --gpus all -p 80:80 emoralesv/ml