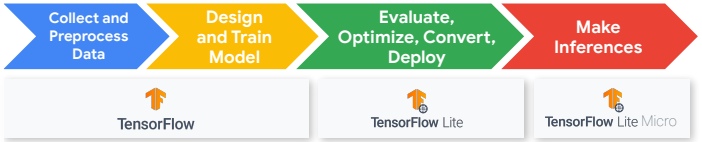
Anomaly Detection Datasets

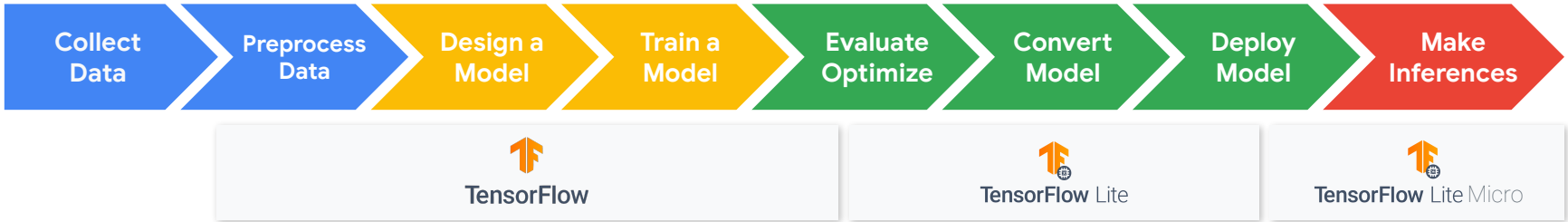
What are we going to learn?

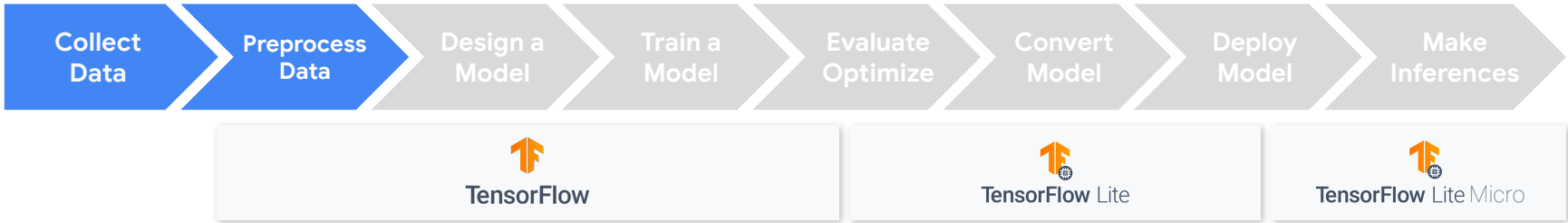
Challenges with an Anomaly Detection Application



Anomaly Detection **ML Pipeline**

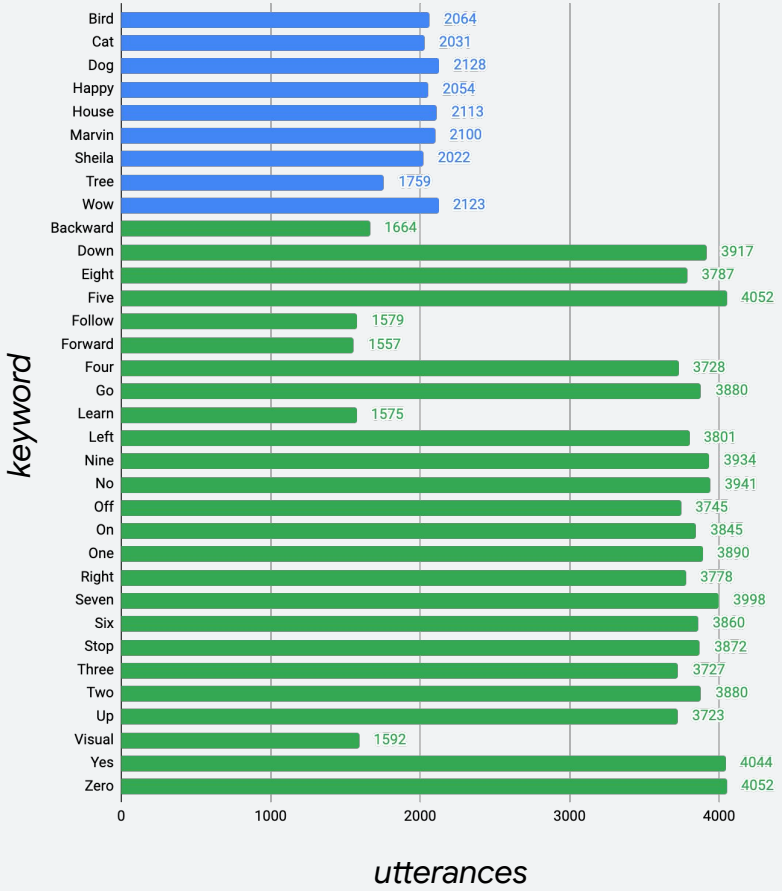
Training, Testing in Colab


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro

The Speech Commands Dataset

- **25 “IoT keywords” +**
  10 “unknown words”
- The set of words is **known in advance**

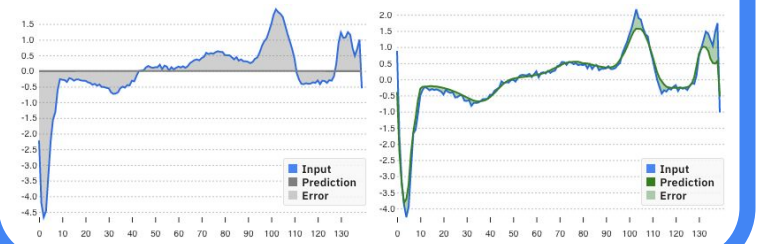


Creating an Anomaly Detection Dataset

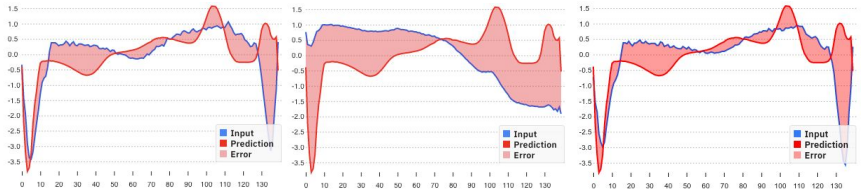
Machine Failure

How are *Anomaly Detection Datasets* **different**?

Normal Data

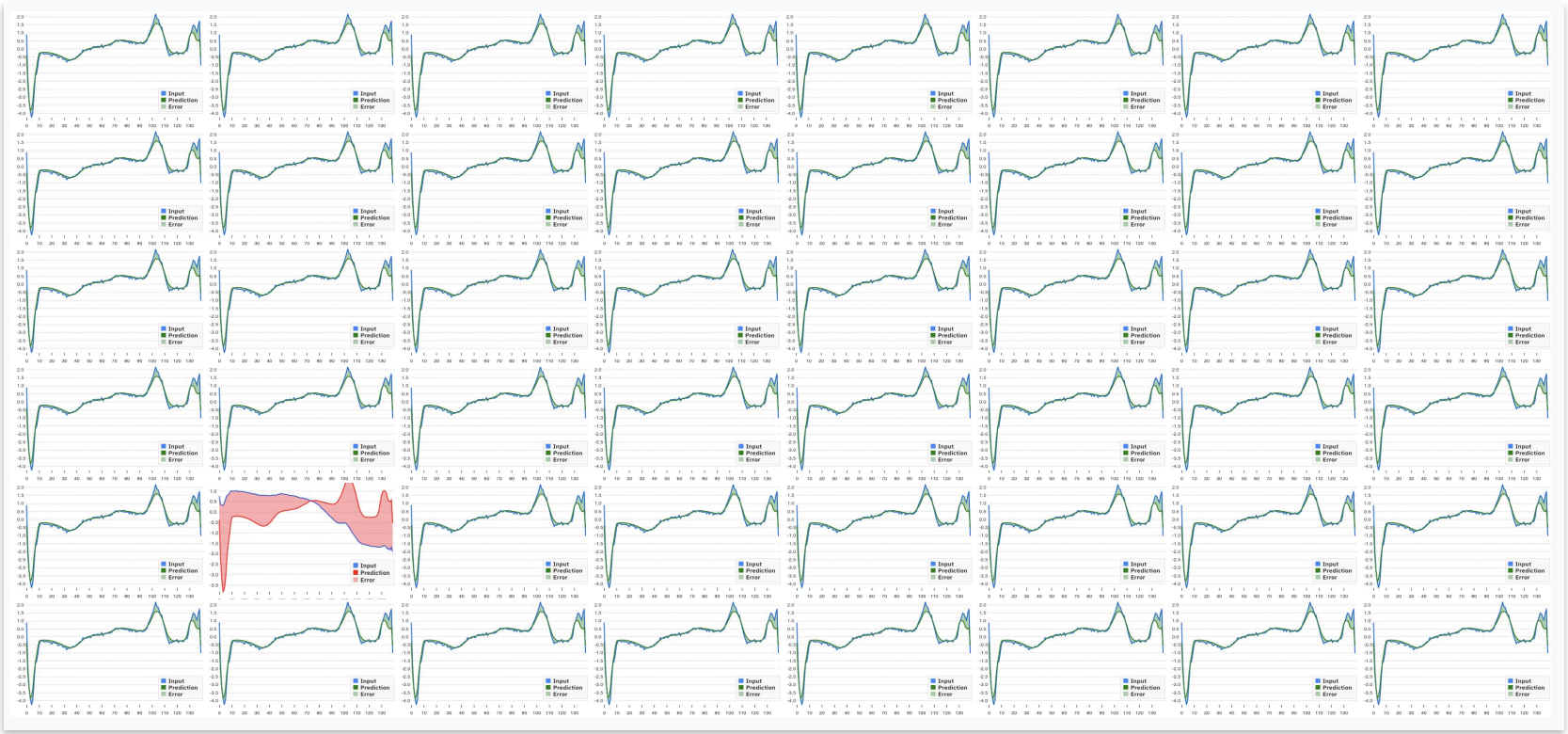


Abnormal Data



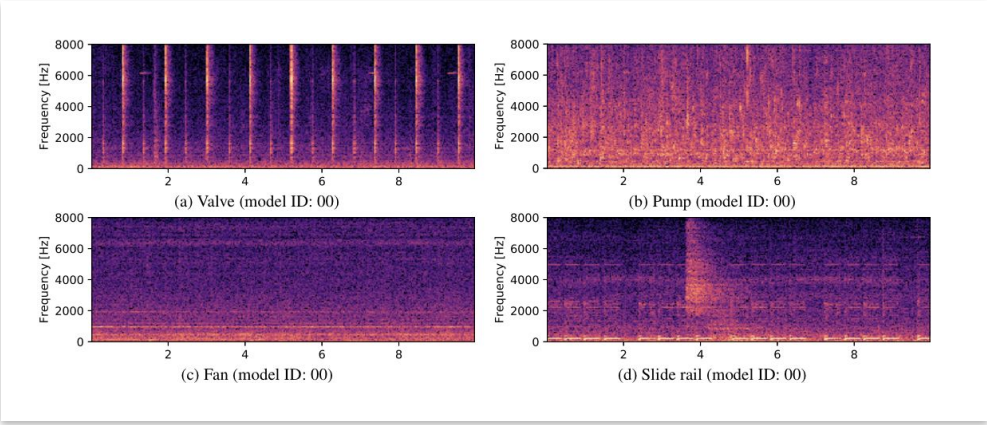
...

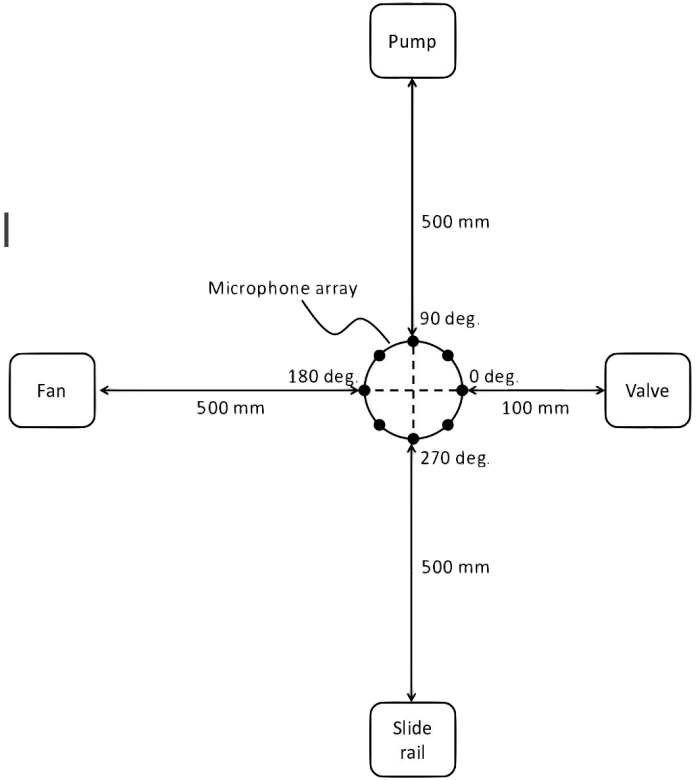
Unbalanced Data



Example: **MIMII Dataset**

- **Sound** Dataset for Malfunctioning Industrial Machine Investigation (MIMII)
- **Anomaly Detection**

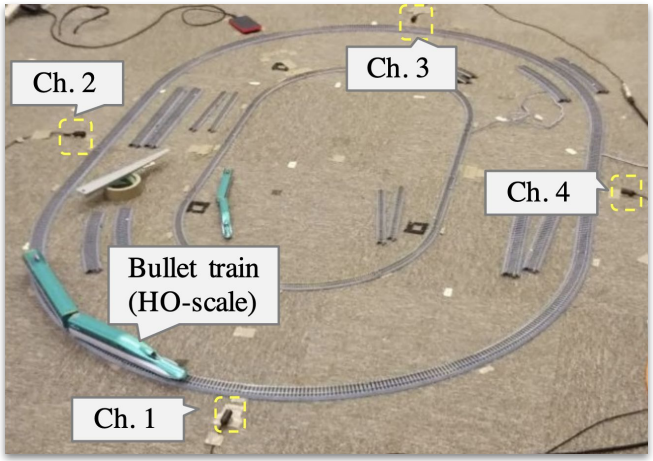


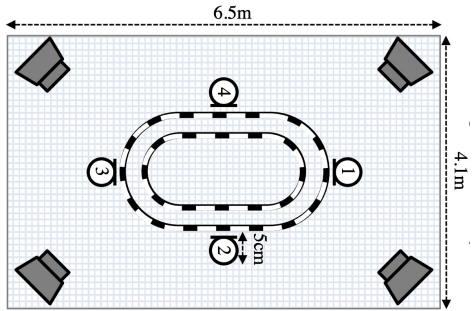


Purohit, Harsh, et al. "MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection." (2019).

Example: **ToyADMOS**

- **540 hours** of sound for *anomaly detection*
- Based on **toys**:
  - Car
  - Conveyor
  - Train
- Anomalies captured by **deliberately damaging** the toys





ToyADMOS Dataset **Characteristics**

- Designed for **three ADMOS tasks**:
  - Product **inspection** (toy car),
  - **Fault diagnosis** for a **fixed** machine (toy conveyor)
  - **Fault diagnosis** for a **moving** machine (toy train)

ToyADMOS Dataset **Characteristics**

- Designed for **three ADMOS tasks**:
  - Product **inspection** (toy car),
  - **Fault diagnosis** for a **fixed** machine (toy conveyor)
  - **Fault diagnosis** for a **moving** machine (toy train)
- ***Machine-operating sounds*** and ***environmental noise*** are individually recorded for simulating various noise levels

ToyADMOS Dataset **Characteristics**

- Designed for **three ADMOS tasks**:
  - Product **inspection** (toy car),
  - **Fault diagnosis** for a **fixed** machine (toy conveyor)
  - **Fault diagnosis** for a **moving** machine (toy train)
- ***Machine-operating sounds*** and ***environmental noise*** are individually recorded for simulating various noise levels
- All sounds are recorded with **four microphones**
  - Testing **noise reduction**
  - **Data-augmentation** techniques such as mix-up

ToyADMOS Dataset **Characteristics**

- Record **multiple machines of same class**
  - Different structure → **Individual variations** despite same toy class

ToyADMOS Dataset **Characteristics**

- Record **multiple machines of same class**
  - Different structure → **Individual variations** despite same toy class
- Record **several times**
  - **Anomalous sound** characteristics from few samples

ToyADMOS Dataset **Characteristics**

- Record **multiple machines of same class**
  - Different structure → **Individual variations** despite same toy class
- Record **several times**
  - **Anomalous sound** characteristics from few samples
- **180 hours** of **normal machine-operating sounds**
  alongside **over 4,000 samples of anomalous sounds**
  - collected at a **48kHz** sampling rate for each task

Low Transferability

- An anomaly detection model will be **very specific** to the training set and therefore **difficult to generalize** to other deployment environments

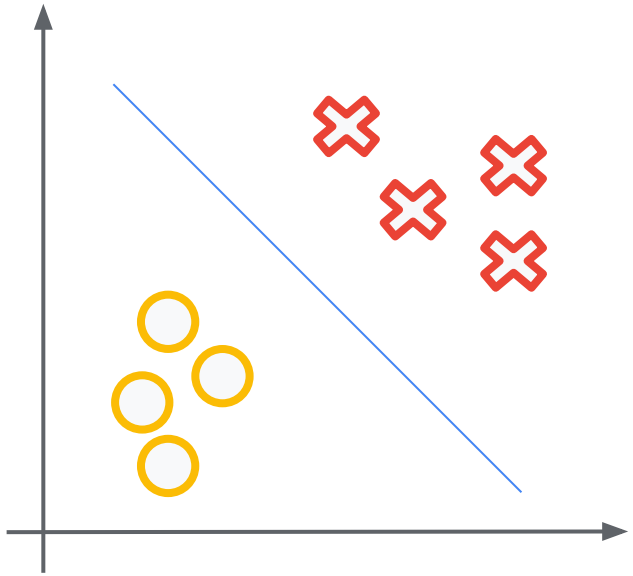
What are we going to learn?

Challenges with an Anomaly Detection Application

Anomaly Detection **ML Pipeline**

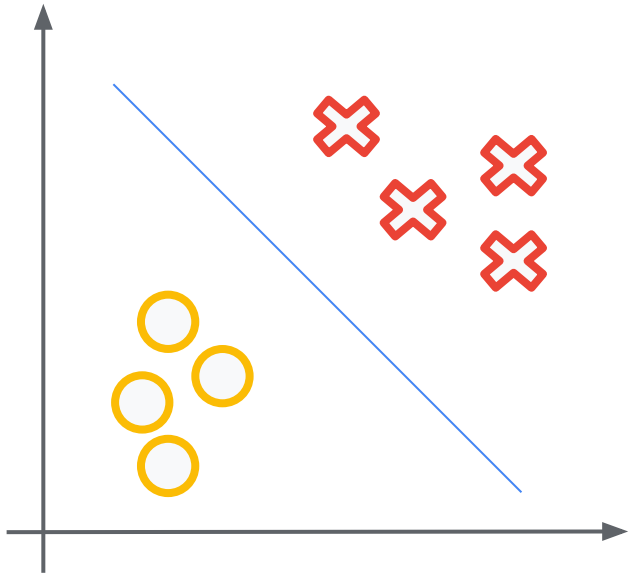
Training, Testing in Colab

Unsupervised learning

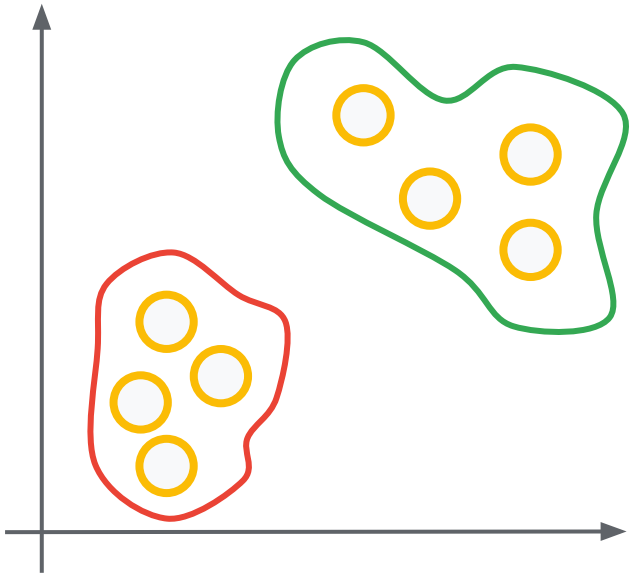


Supervised Learning

Unsupervised learning



Supervised Learning



Unsupervised Learning