



MACHINE LEARNING FOR DATA QUALITY MONITORING (DQM) AT CMS

PHYSICS RESEARCH SEMESTER ABROAD

MENTOR | FEDERICO DE GUIO, Ph.D.

STUDENT | GUILLERMO A. FIDALGO RODRÍGUEZ

CMS DETECTOR

Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel ($100 \times 150 \mu\text{m}$) $\sim 16\text{m}^2 \sim 66\text{M}$ channels
Microstrips ($80 \times 180 \mu\text{m}$) $\sim 200\text{m}^2 \sim 9.6\text{M}$ channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying $\sim 18,000\text{A}$

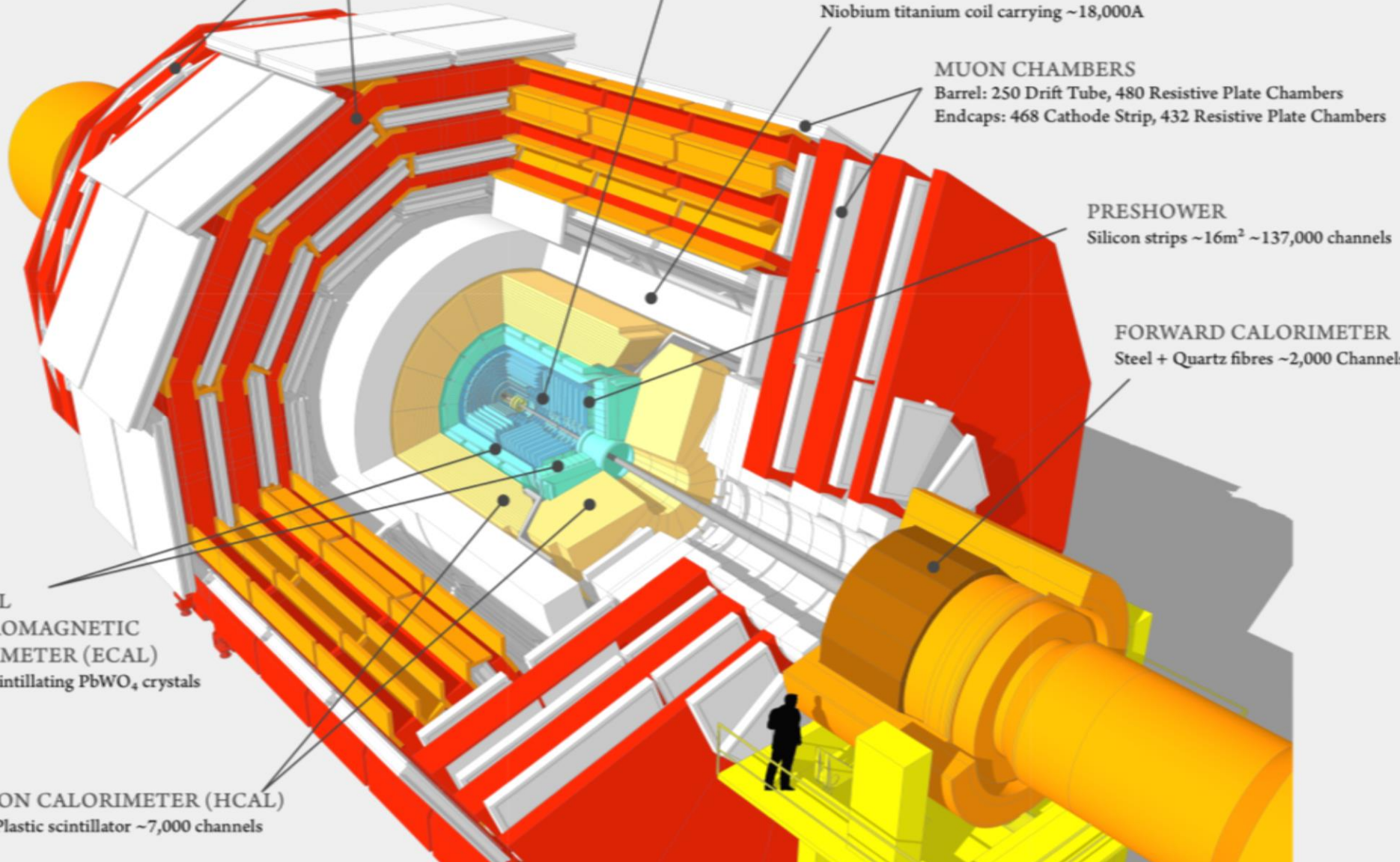
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers

PRESHOWER
Silicon strips $\sim 16\text{m}^2 \sim 137,000$ channels

FORWARD CALORIMETER
Steel + Quartz fibres $\sim 2,000$ Channels

CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
 $\sim 76,000$ scintillating PbWO_4 crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator $\sim 7,000$ channels



THE COMPACT MUON SOLENOID (CMS) EXPERIMENT

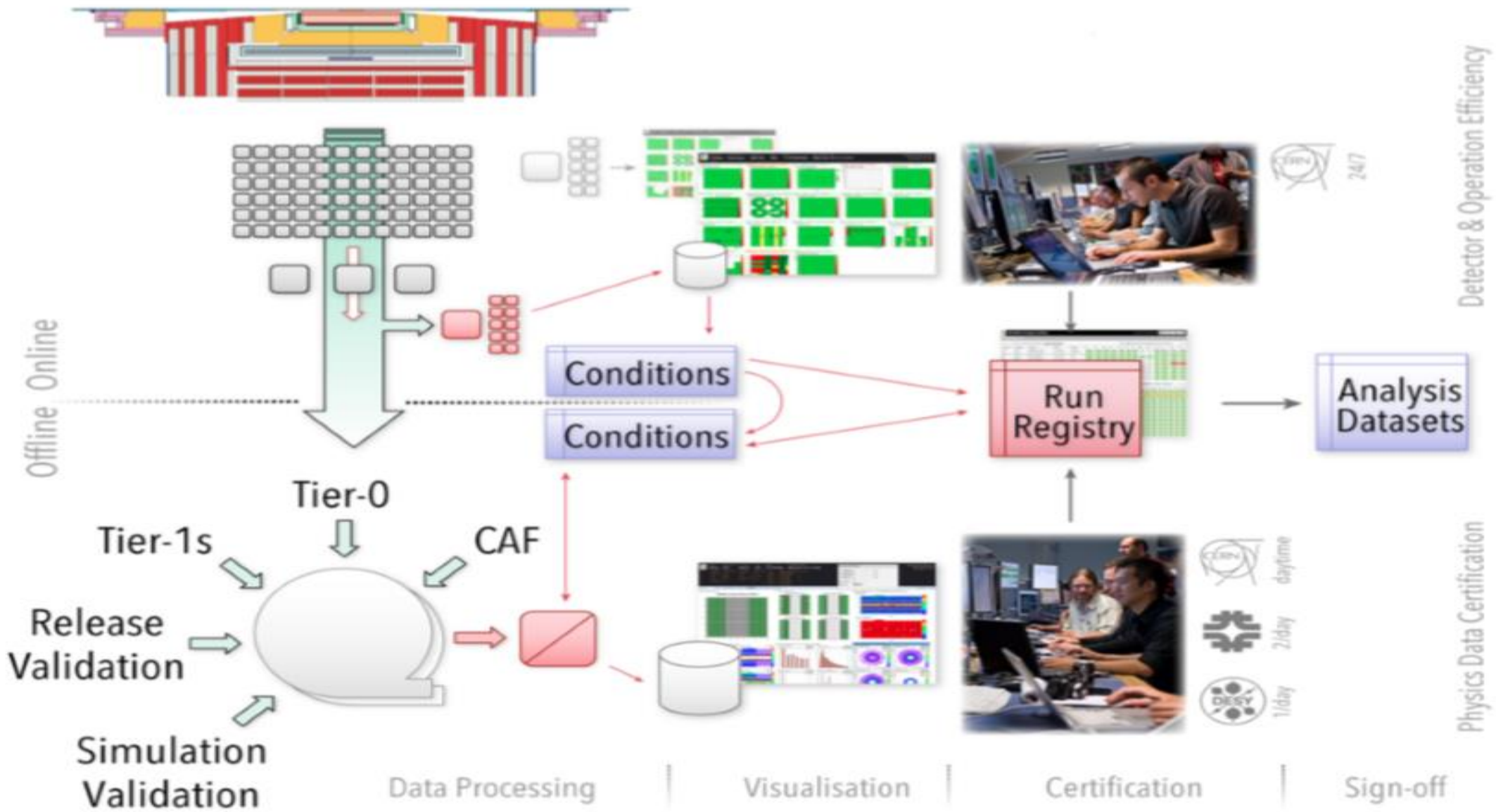
<http://cms.web.cern.ch/news/what-cms>

THE CHALLENGE

- You have to make sure that it behaves well in order to perform sensible data analysis.
- Reduce man power.
 - Shifters monitor constantly the quality of the data flow.
 - Discriminate between good and bad data to have high purity
 - Build something that helps the people to minimize the time needed to spot problems and save time examining hundreds of histograms
 - Build intelligence that analyzes the data and raises alarms in case of problems. Have quick feedback.

WHAT IS DATA QUALITY MONITORING (DQM)?

- 2 workflows:
- Online DQM
 - Provides feedback of live data taking.
 - Alarms if something goes wrong.
- Offline DQM
 - After data taking
 - It is responsible of bookkeeping and certifying the final data with fine time granularity.



HOW TO AUTOMATIZE THE DATA QUALITY CHECKS? USE MACHINE LEARNING!

- It's everywhere now!
 - A.I. Learning
 - Self-driving cars
 - How does Google/Facebook know what you want?
 - Face/Handwriting Recognition
- In our case everything reduces to a Classification problem
 - Anomaly Detection

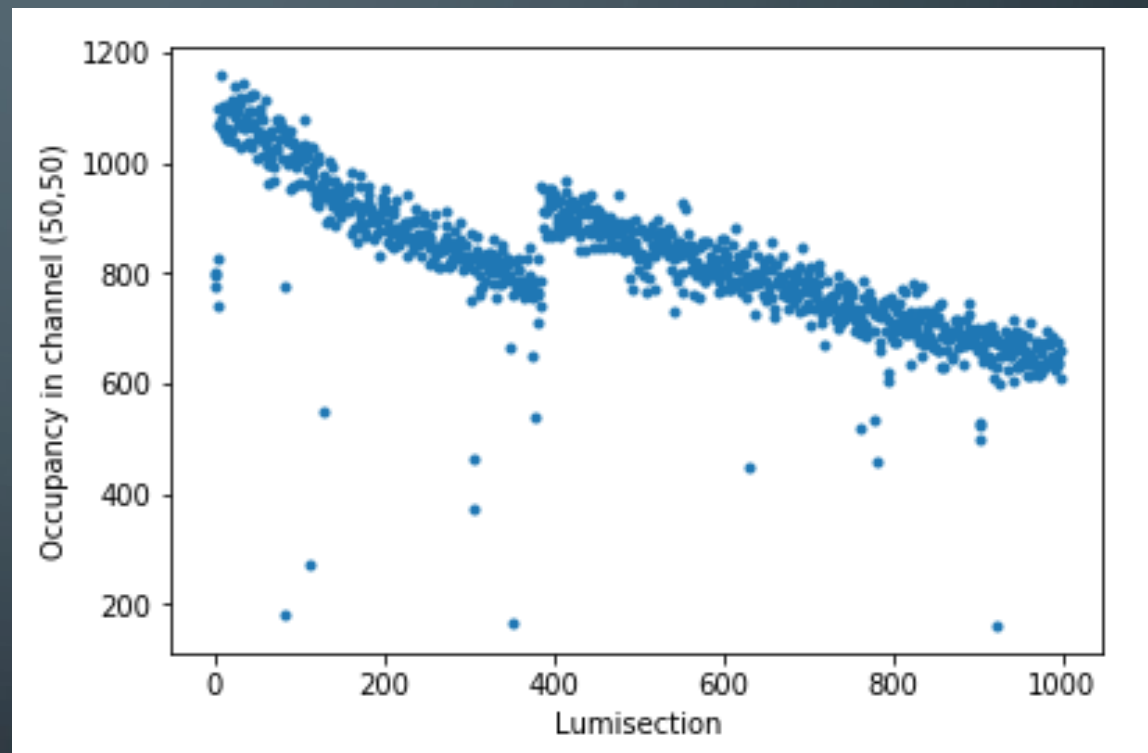
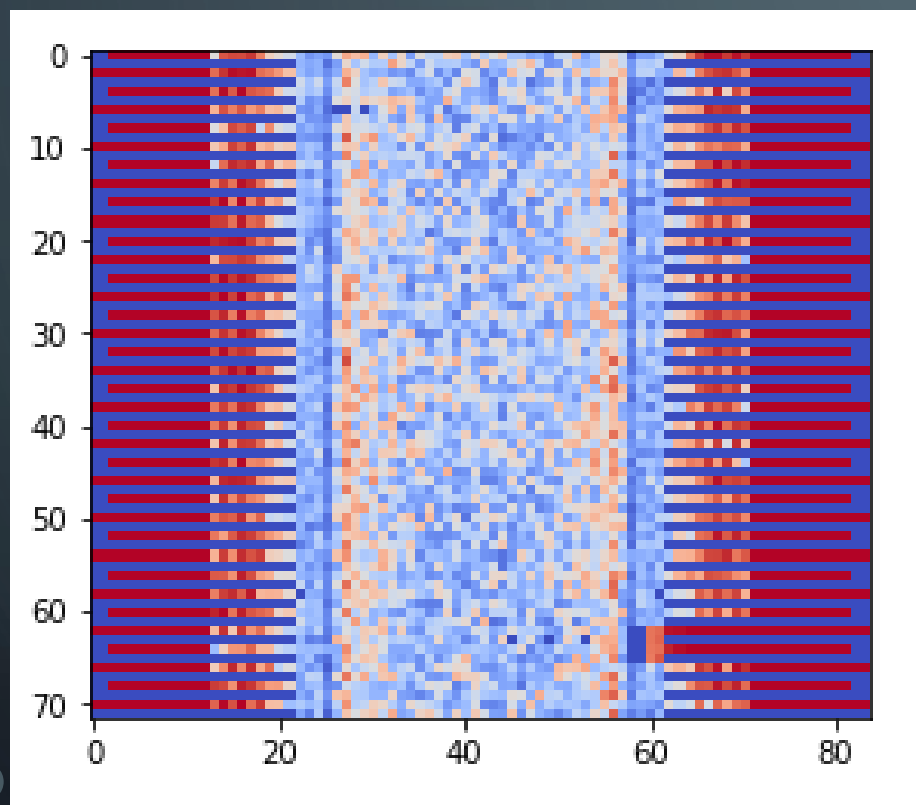
OBJECTIVES AND MY CONTRIBUTION

- The project aims at applying recent progress in Machine Learning techniques to the automation of the DQM scrutiny for HCAL
 - Focus on the Online DQM.
 - Compare the performance of different ML algorithms.
 - Fully supervised vs unsupervised approach.

TOOLS AND DATA PREPARATION

- Have been familiarized with the following tools:
 - Working with data stored as hdf5 files
 - Familiarize with NumPy arrays
 - Working env: Juypiter python notebook
 - Matplotlib is used for plotting results
- Data comes in form of occupancy maps for HCAL
 - Flow of one map each lumisection for every lumisection.

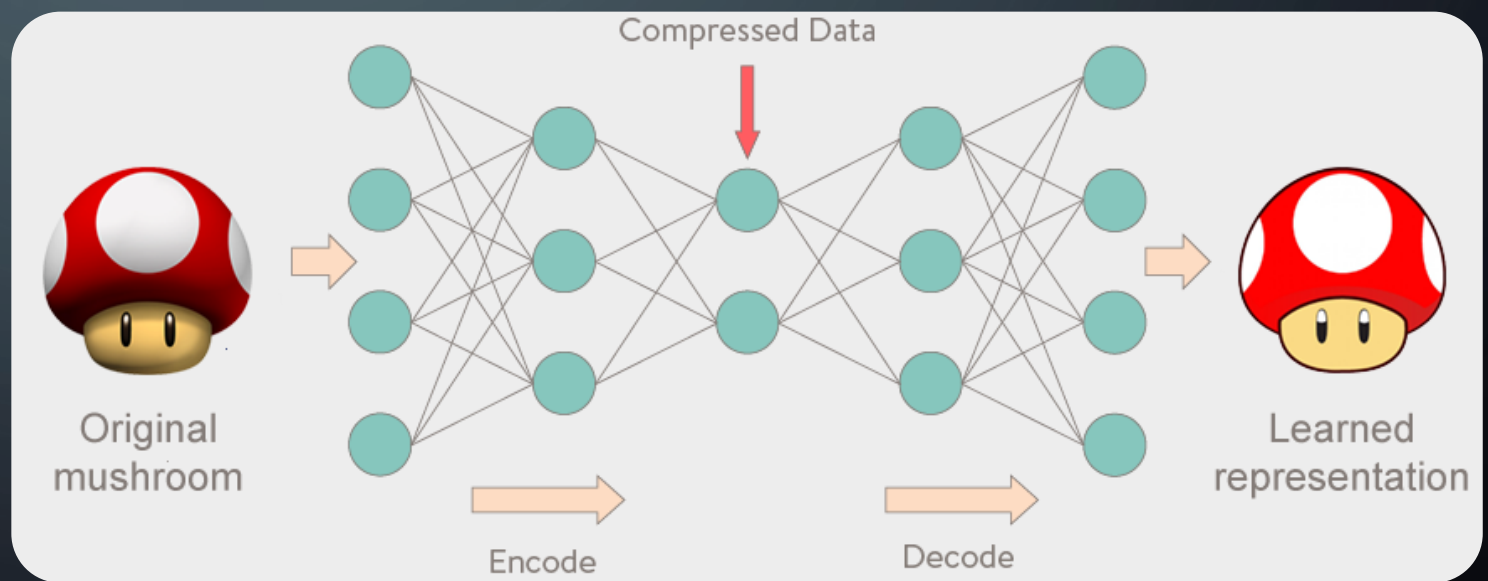
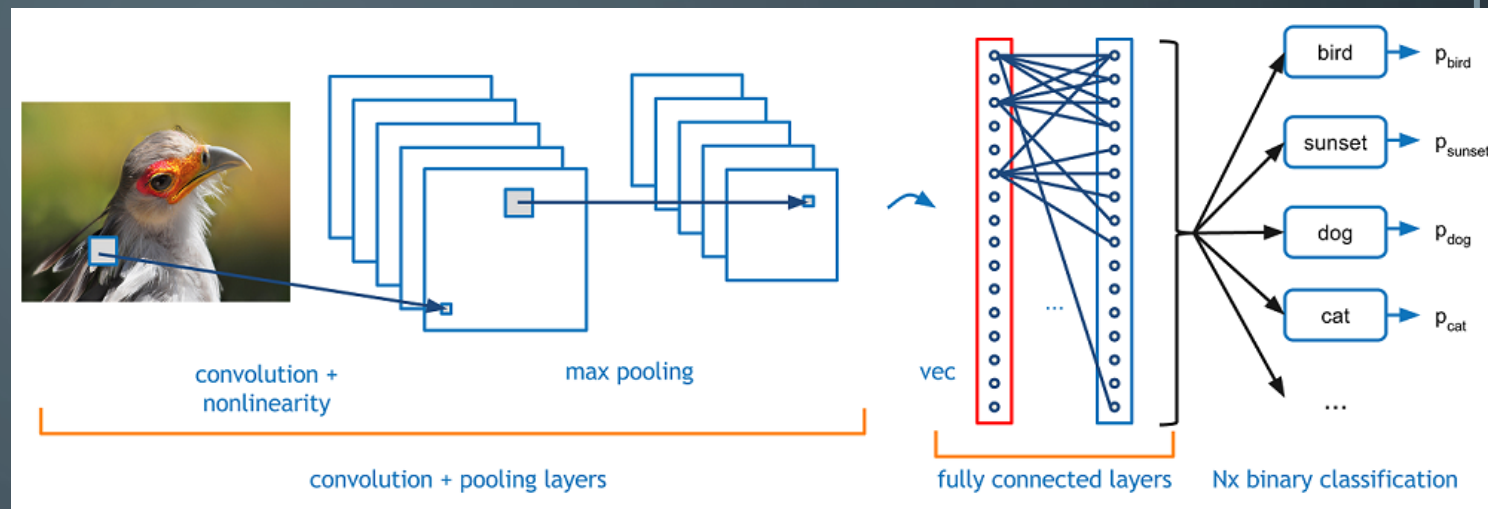
IMAGES



WHAT'S NEXT?

- Familiarize with Keras
 - Creation of a model
 - Train it, test its performance
- Compare it to other models
 - CNN
 - AE

CNN



AE