

Components of a machine learning project

Tom Rochette <tom.rochette@coreteks.org>

March 22, 2020 — [8c4db961](#)

0.1 Context

0.2 Learned in this study

0.3 Things to explore

1 Overview

- Data: The units of data from which information/knowledge will be extracted (text, csv, binary, json, databases)
- Extract, Transform, Load (ETL)/Feature extractor pipeline: A set of tools/procedure to convert the data into a format more appropriate for further processing, as well as storing (google refine)
- Model/Architecture: The machine learning model that converts incoming data into predictions/decisions (tensorflow, keras, torch)
- API: HTTP endpoint that can be easily called to pass information to the system or query the model for information (flask)
 - Request data from data source
 - Run inference given data
- Benchmark: Test changes to your model to see how they fare against previous models
- Containers: Encapsulates the tools and procedures used to process data and turn it into information. The container makes it easy to install dependencies and to reproduce the environment on other machines (docker)
- Container-orchestrator: Takes care of running containers according to certain policies, making it possible to scale a project if required (kubernetes, helm)

2 See also

3 References

- <https://www.infoworld.com/article/3198252/machine-learning/data-in-intelligence-out-machine-learning-pipelines-demystified.html>