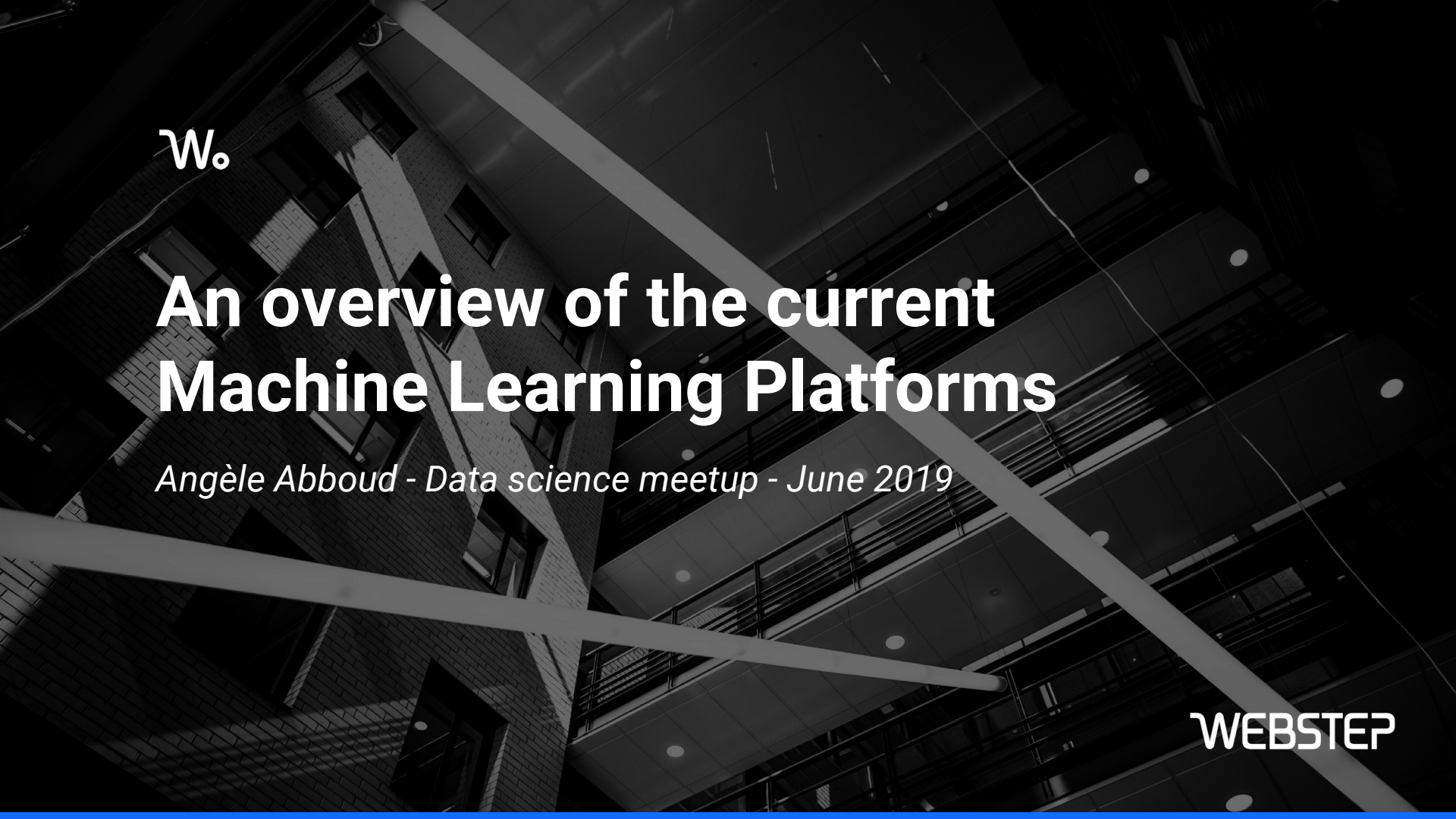
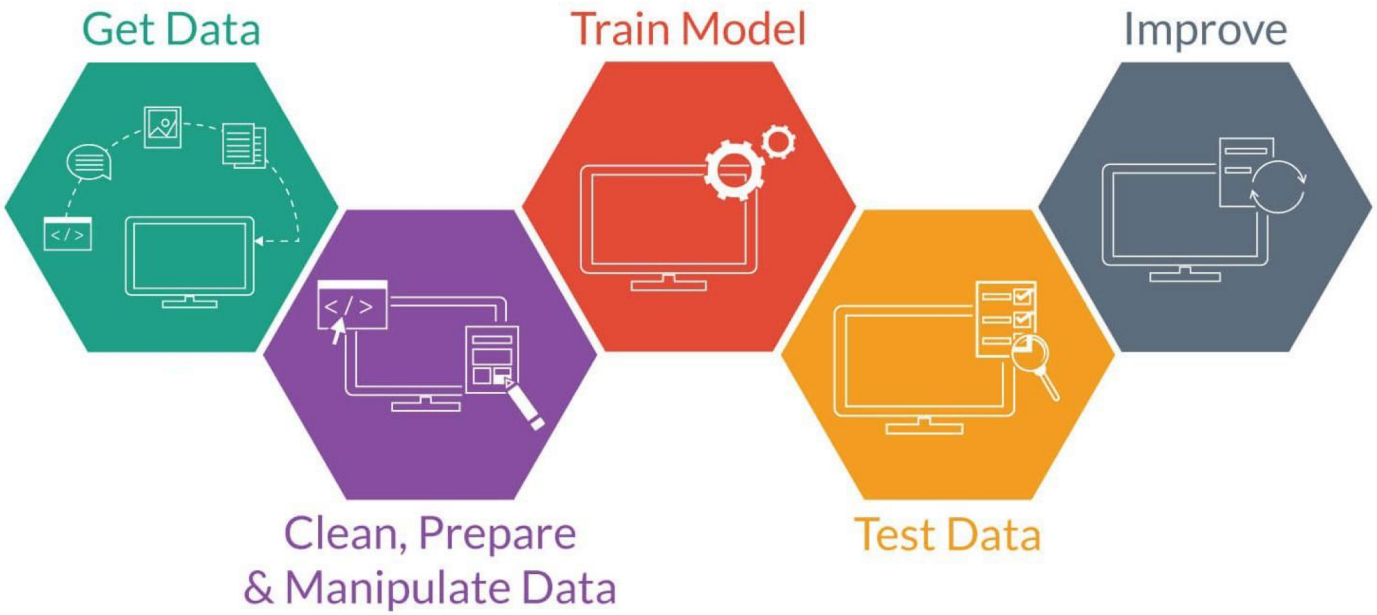


An overview of the current Machine Learning Platforms

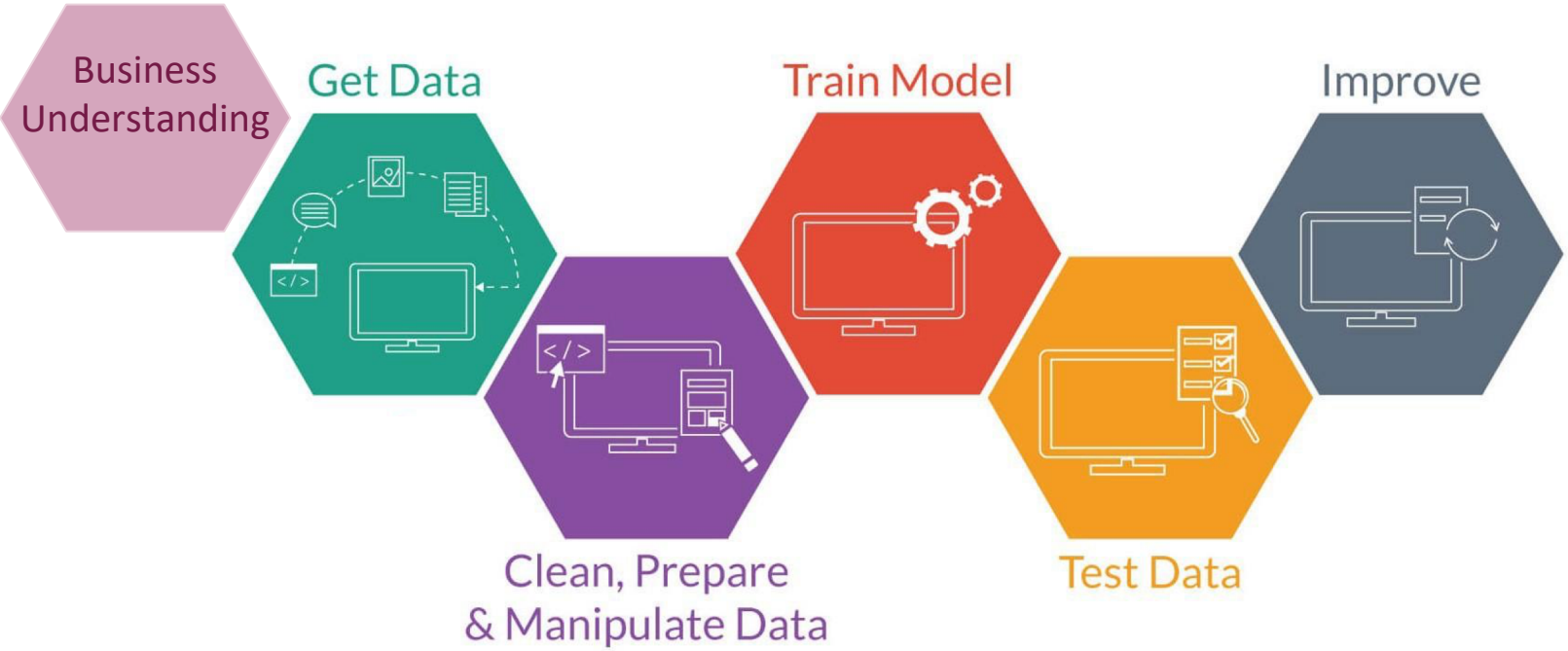
Angèle Abboud - Data science meetup - June 2019

WEBSTEP

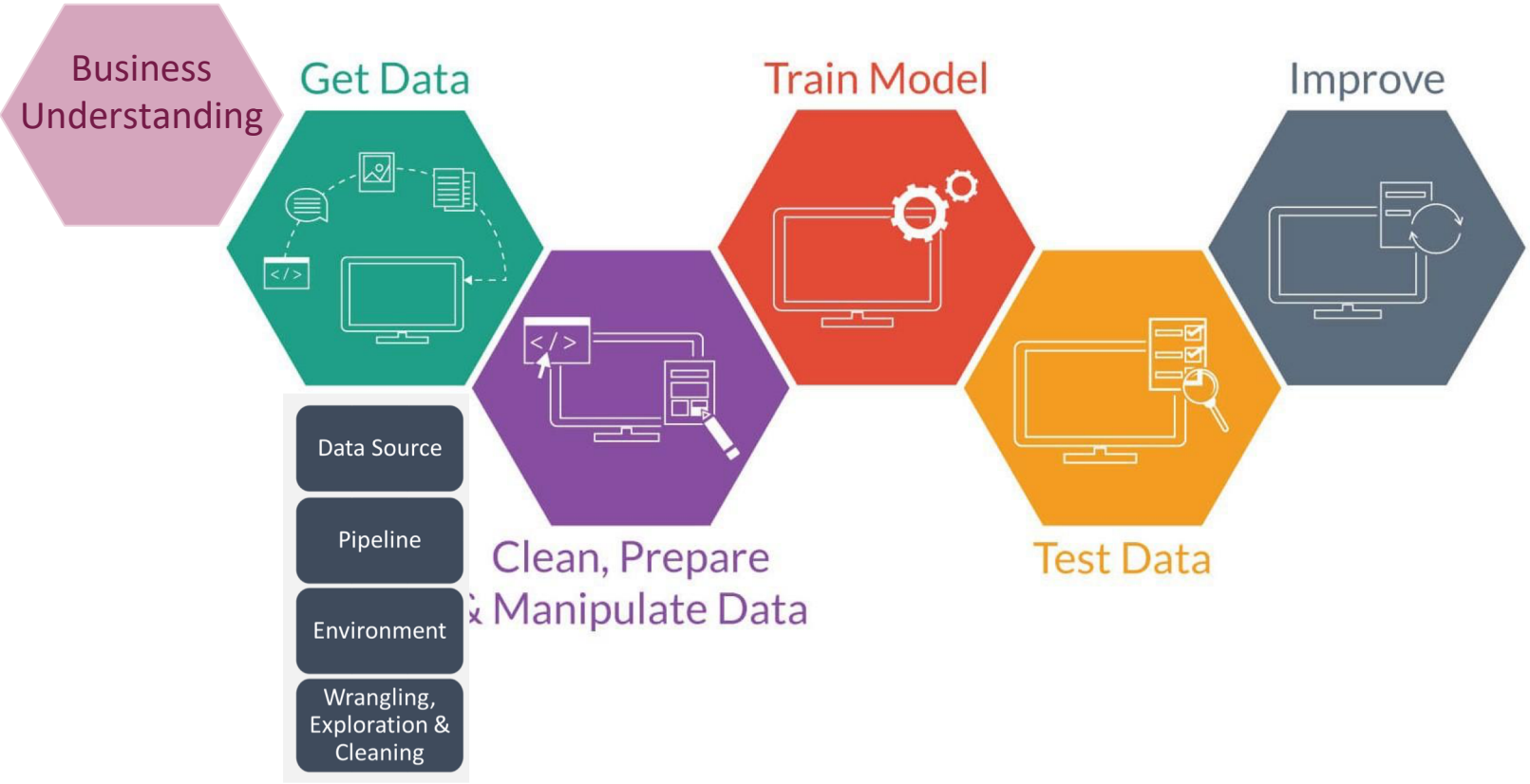
Machine Learning Workflow



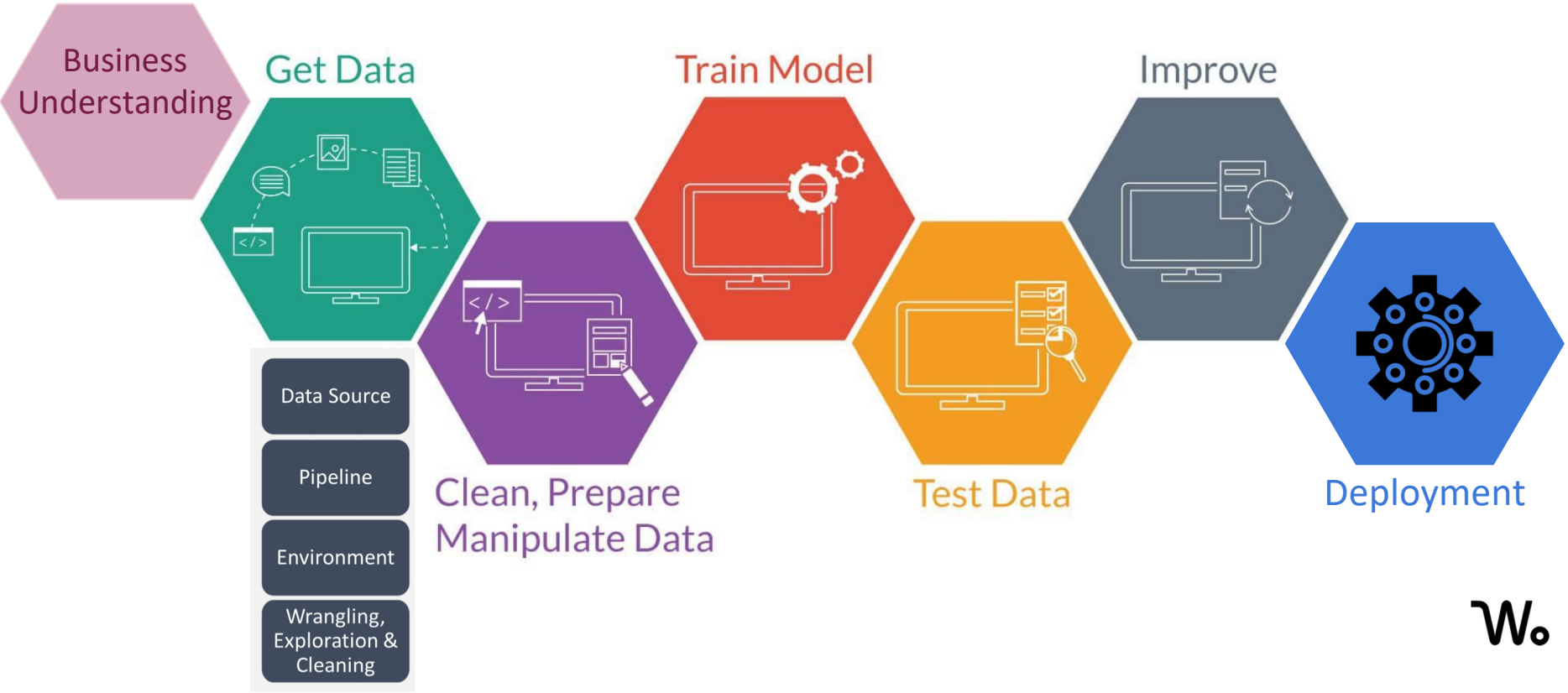
Machine Learning Workflow



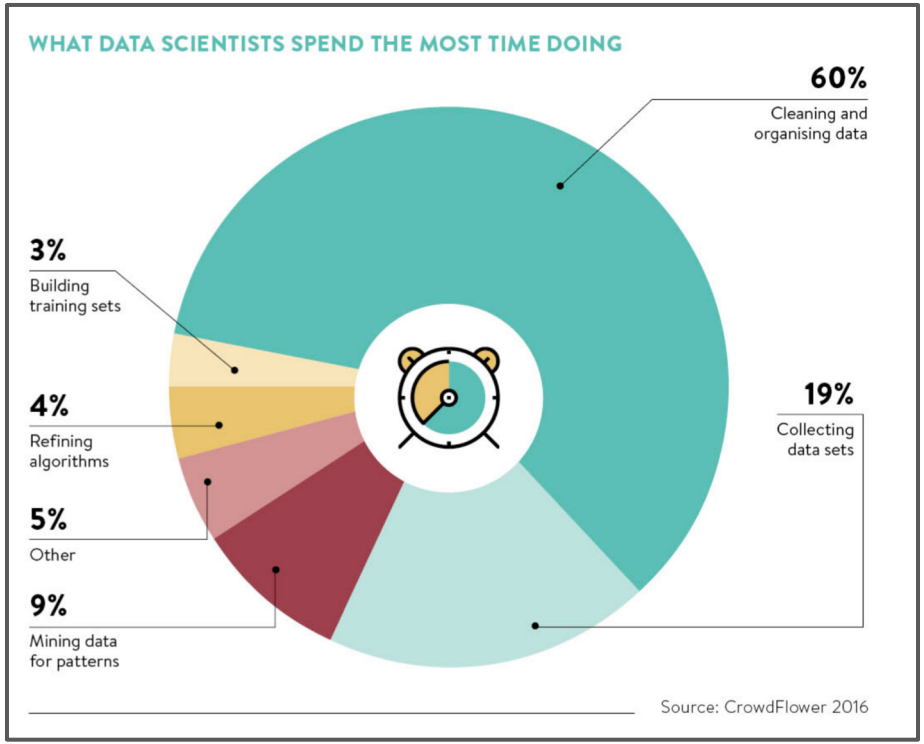
Machine Learning Workflow

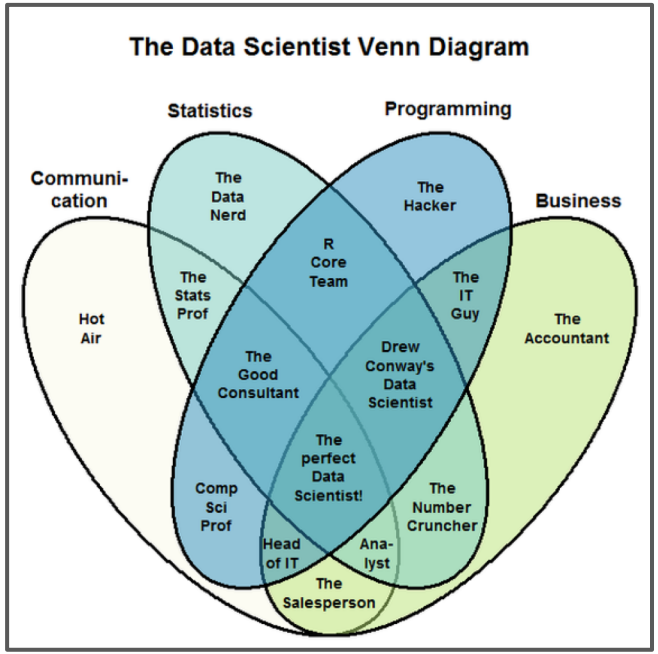


Machine Learning Workflow

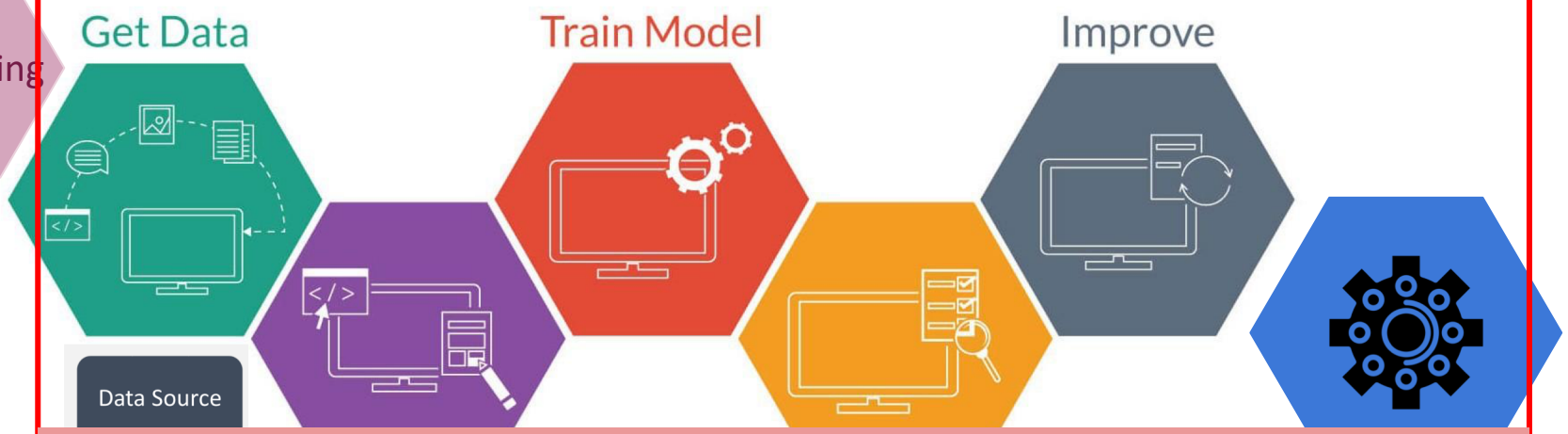


Data scientist Workflow





Business Understanding



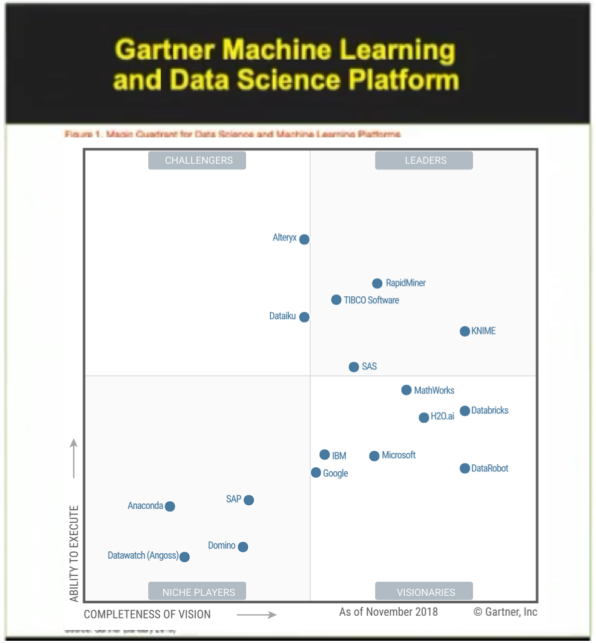
- Automation Machine Learning -
Building technologies and tools around this workflow to be more productive

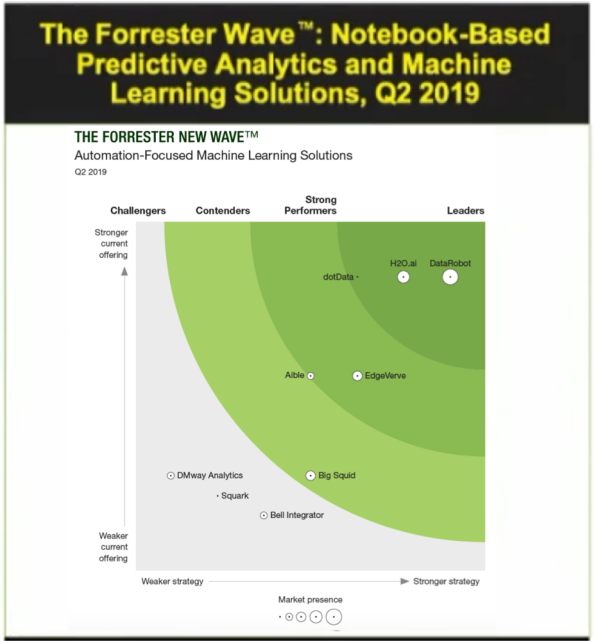
- Machine Learning Platforms -
To make data science accessible to field expert

Machine Learning and Data Science Platforms

Where to start? Don't google!

"It's well rounded, and the vision and roadmap are better than most competitors"





An non exhaustive list...

Entreprise ML Platforms	Open source/Free ML Platforms
- RapidMiner - TIBCO Software - **KNIME** - SAS - Dataiku - **H20 Driverless AI** - **Datarobot** - **Microsoft Azure ML** - Google cloud autoML - Alteryx - Databricks - IBM Watson	- **Anaconda** - **H20** - Dataiku Lite Edition - **KNIME analytics Platform** AutoML tools: - auto-sklearn - auto-weka - tpot...

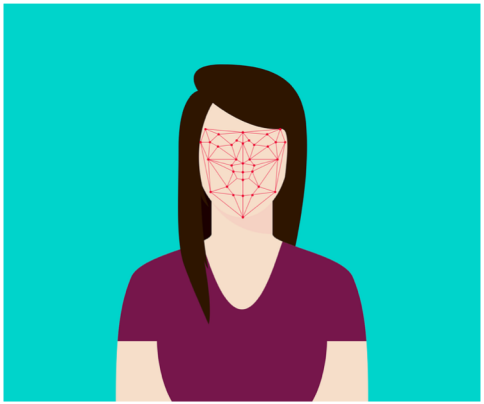
10 criteria to go through the list… again

1. Data preparation features
2. Bank of pre-written algorithms
3. Scalability
4. Support for open source programming languages
5. A workbench-style interface
6. Collaboration features that allow project sharing
7. Deployment capabilities
8. Model management tools to track the effectiveness of models in production
9. Pre-written tools for common business problems (customer churn modeling)
10. The ability of the vendor to execute on their promises.
11. BUDGET

https://www.gartner.com/reviews/market/data-science-machine-learning-platforms

https://www.gartner.com/reviews/market/data-science-machine-learning-platforms/compare/product/alteryx-promote-vs-datarobot-vs-ibm-watson-studio?utm_source=gartner&utm_medium=email&utm_campaign=share&utm_content=mail_icon

How does it look in real?



An non exhaustive list...

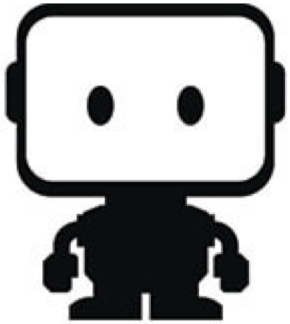
Entreprise ML Platforms	Open source/Free ML Platforms
- RapidMiner - TIBCO Software - **KNIME** - SAS - Dataiku - **H20 Driverless AI** - **Datarobot** - **Microsoft Azure ML** - Google cloud autoML - Alteryx - Databricks - IBM Watson	- **Anaconda** - **H20** - Dataiku Lite Edition - **KNIME analytics Platform** AutoML tools: - auto-sklearn - auto-weka - tpot...

Entreprise ML Platforms

How DataRobot works

1. Ingest your data
2. Select the target variable
3. Build hundreds of models in one click
4. Explore top models and get insights
5. Deploy the best model

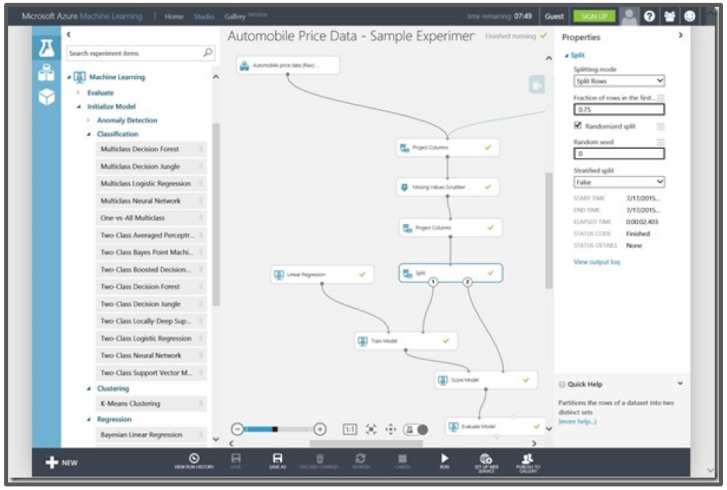
Example1: DataRobot - Citizen Data scientist



DataRobot

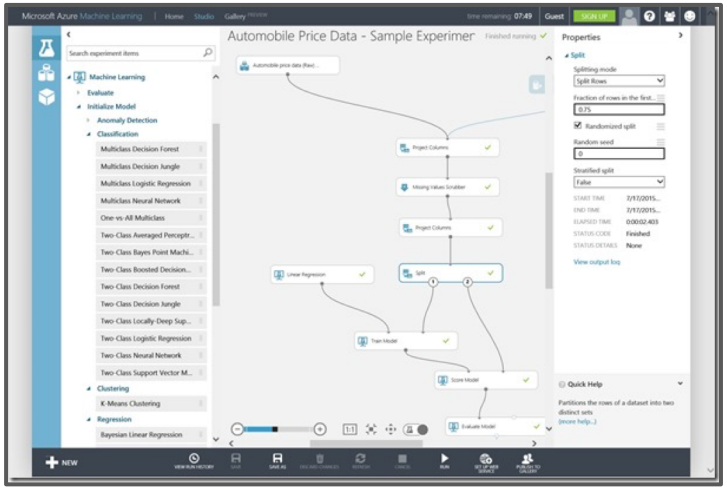
- Leader in augmented ML = Great interface
- Entreprise driven and not really for expert Data Scientist
- Expensive
- Not much flexibility

Microsoft - Azure ML Studio



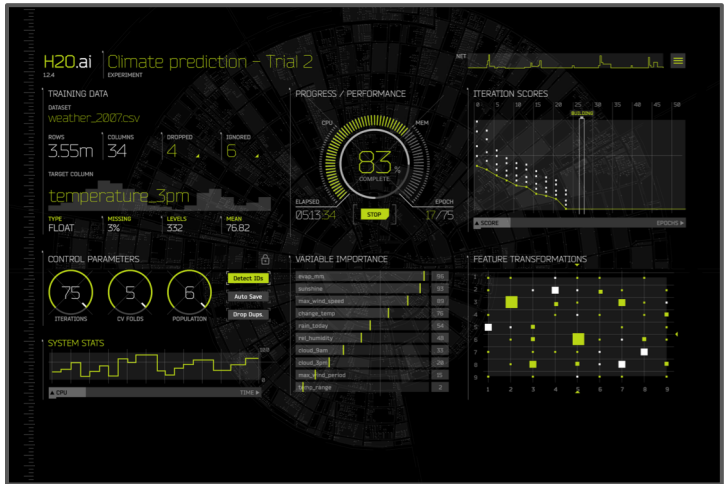
Cloud infrastructure approach
Great variety of components but complex in the offer
High flexibility

Microsoft - Azure ML Studio



Cloud infrastructure approach
Great variety of components but complex in the offer
High flexibility

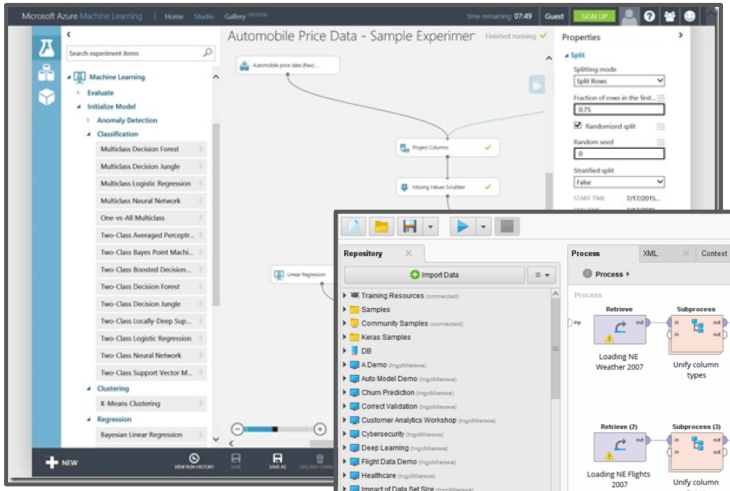
H20 driverless AI

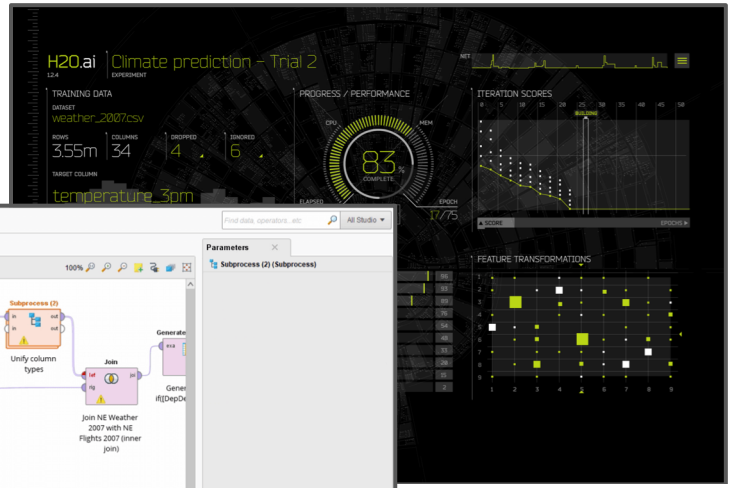


Industry started in ML performance
Deep-Learning and GPU Layer
Important automation capabilities
Highly scalable

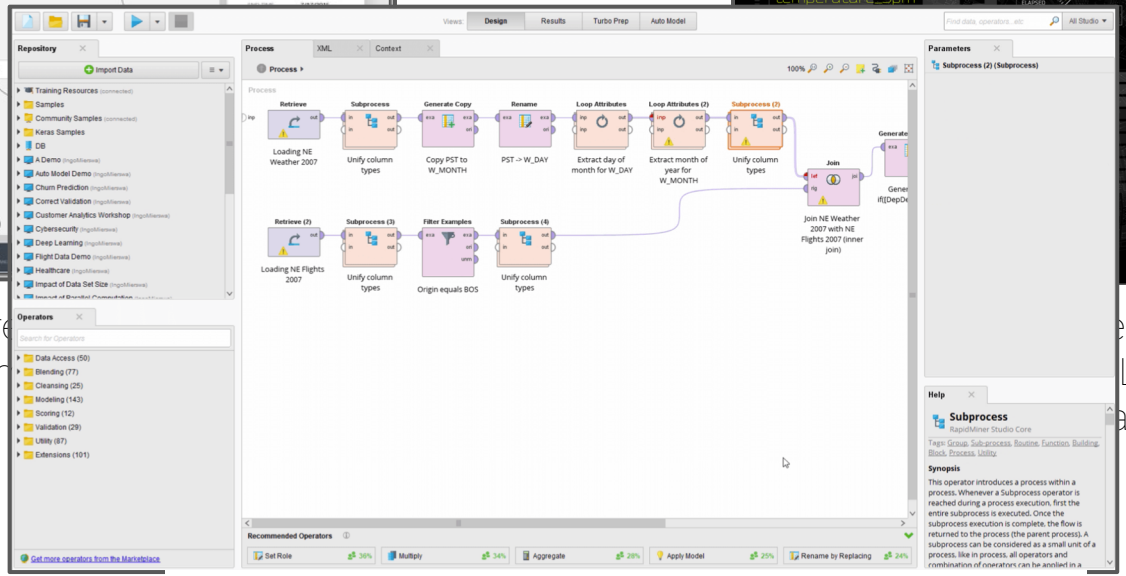
Microsoft - Azure ML Studio

H20 driverless AI





RapidMiner



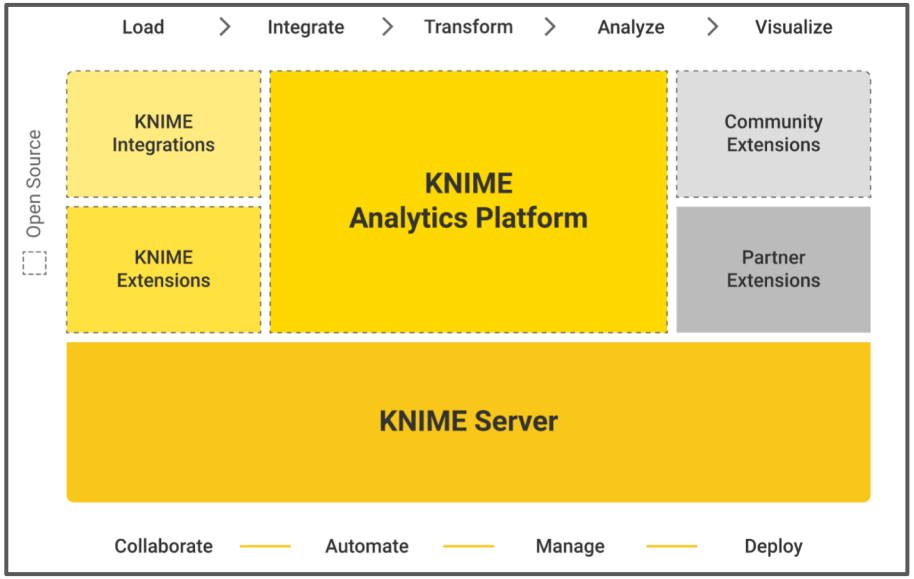
Cloud infrastructure
Great variety of con
High flexibility

erformance
Layer
apabilities

From citizen Data scientist to Expert
Coherent and simple end-to-end platform
Low-key on monitoring
Difficult pricing model

Open-Source ML Platforms

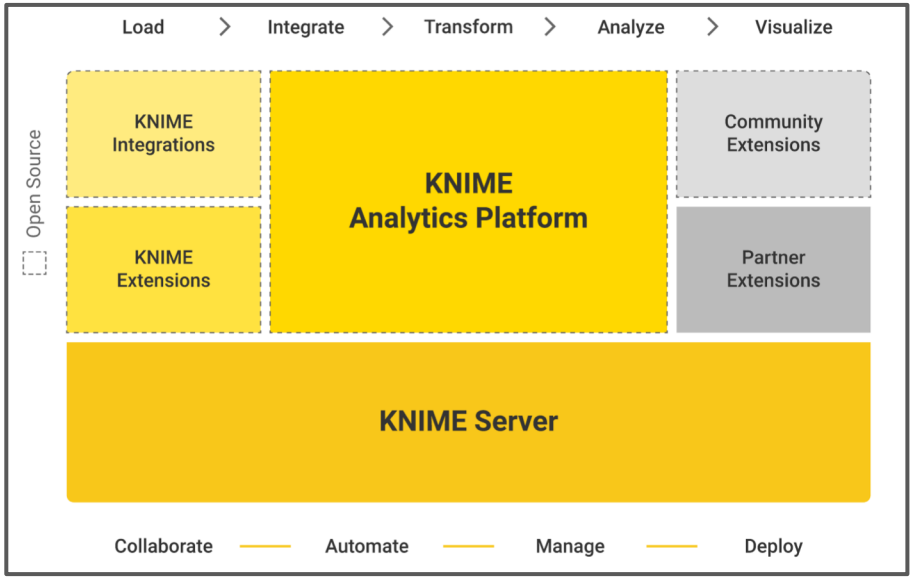
KNIME - Analytics Platform



85% of capabilities are open-source
Knime API
Lack of scalability and visibility

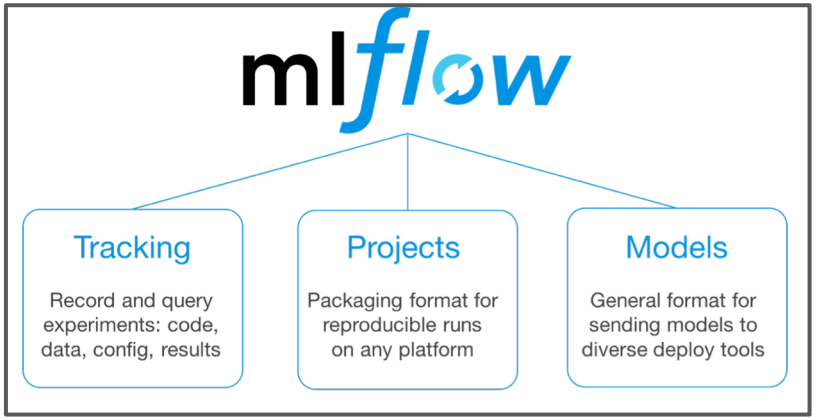
Open-Source ML Platforms

KNIME - Analytics Platform



85% of capabilities are open-source
Knime API
Lack of scalability and visibility

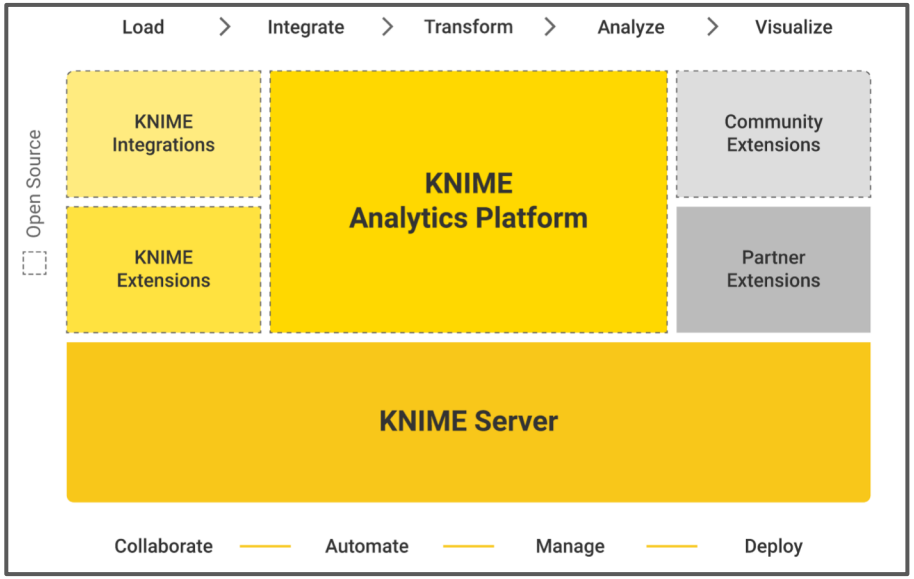
DataBricks MLflow



Framework for end-to-end ML process
Easy deployment with Apache Spark/ Microsoft Azure / or docker container

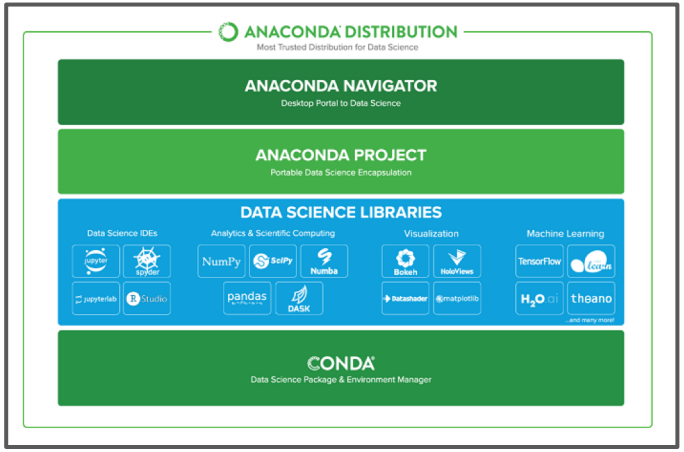
Open-Source ML Platforms

KNIME - Analytics Platform



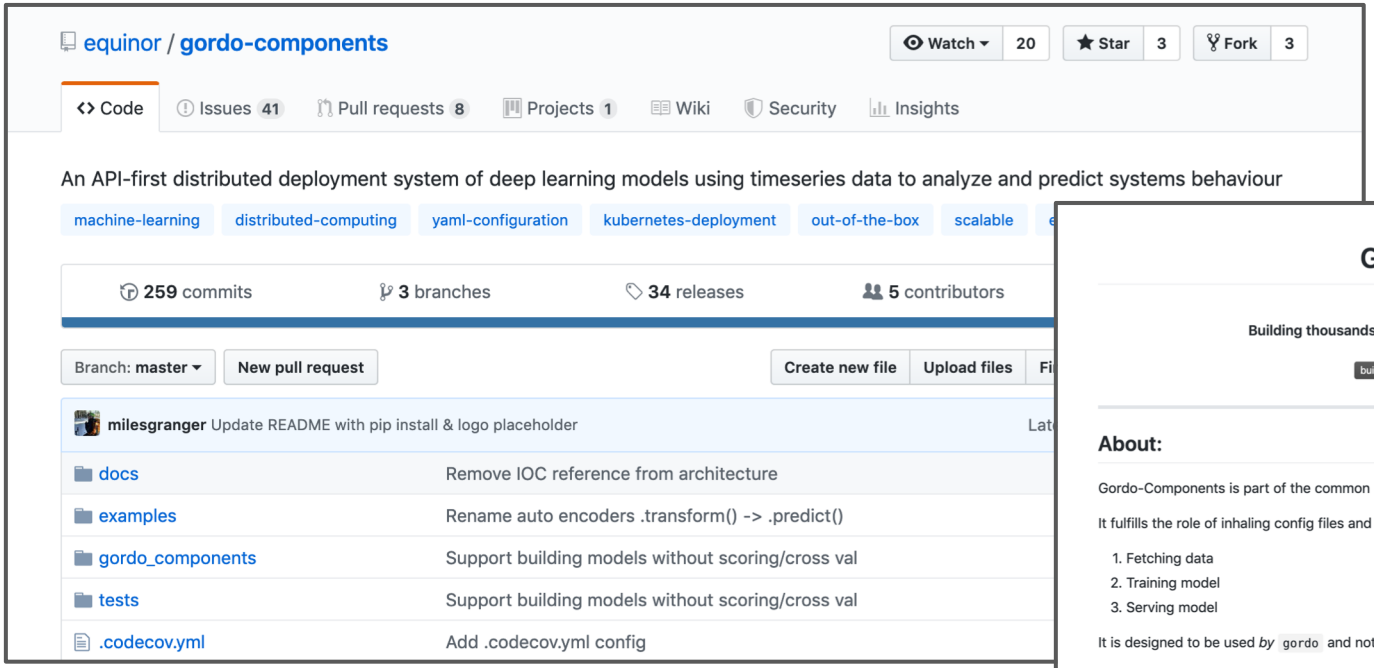
85% of capabilities are open-source
Knime API
Lack of scalability and visibility

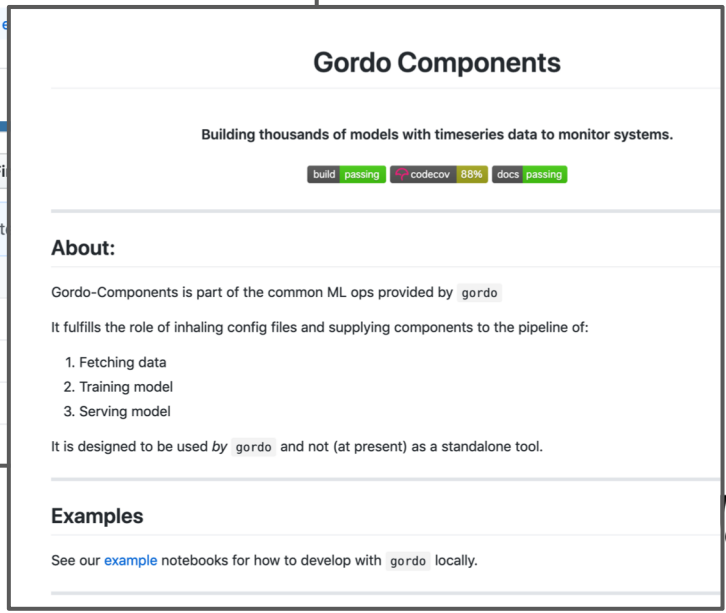
Anaconda



THE data science platform based on python
Wide source of libraries and open-source project
For expert data scientist

Local initiatives: Gordo-components





Local initiatives: Gordo-components

equinor / **gordo-components**

Watch 20 | Star 3 | Fork 3

Code | Issues 41 | Pull requests 8 | Projects 1 | Wiki | Security | Insights

An API-first distributed deployment system of deep learning models using timeseries data to analyze and predict systems behaviour

machine-learning | distributed-computing | yaml-configuration | kubernetes-deployment | out-of-the-box | scalable | end-to-end

259 commits | **3** branches | **34** releases | **5** contributors | AGPL-3.0

Branch: **master** | New pull request | Create new file | Upload files | Find File | Clone or download

milesgranger Update README with pip install & logo placeholder — Latest commit `5fc587b` 6 days ago

docs	Remove IOC reference from architecture	16 days ago
examples	Rename auto encoders .transform() -> .predict()	6 days ago
gordo_components	Support building models without scoring/cross val	23 hours ago
tests	Support building models without scoring/cross val	23 hours ago
.codecov.yml	Add .codecov.yml config	10 days ago

Miles Granger
- Data and Machine learning Engineer -

Natalie Caruana
- Data Scientist and Analyst -

Home Message

1. It is a technology to know about
2. You won't make a big mistake by choosing one
3. You have to choose yourself, but here are you some questions to help:

   - Who will be using it: citizen DS, data engineer, data analyst, expert data scientists….?
   - In which part of the process do you need automation: data prep, deployment, visualization?
   - Collaboration within a team?
   - Do you need multiple users to access the model?
   - What is your data type?
   - What is the budget?