

pharo-ai

Sebastian JORDAN MONTANO¹

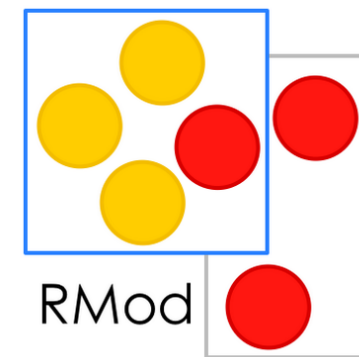
sebastian.jordan@inria.fr

Oleksandr ZAITSEV^{2,1}

oleksandr.zaitsev@arolla.fr

¹ Inria, Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRISTAL

² Arolla, Paris

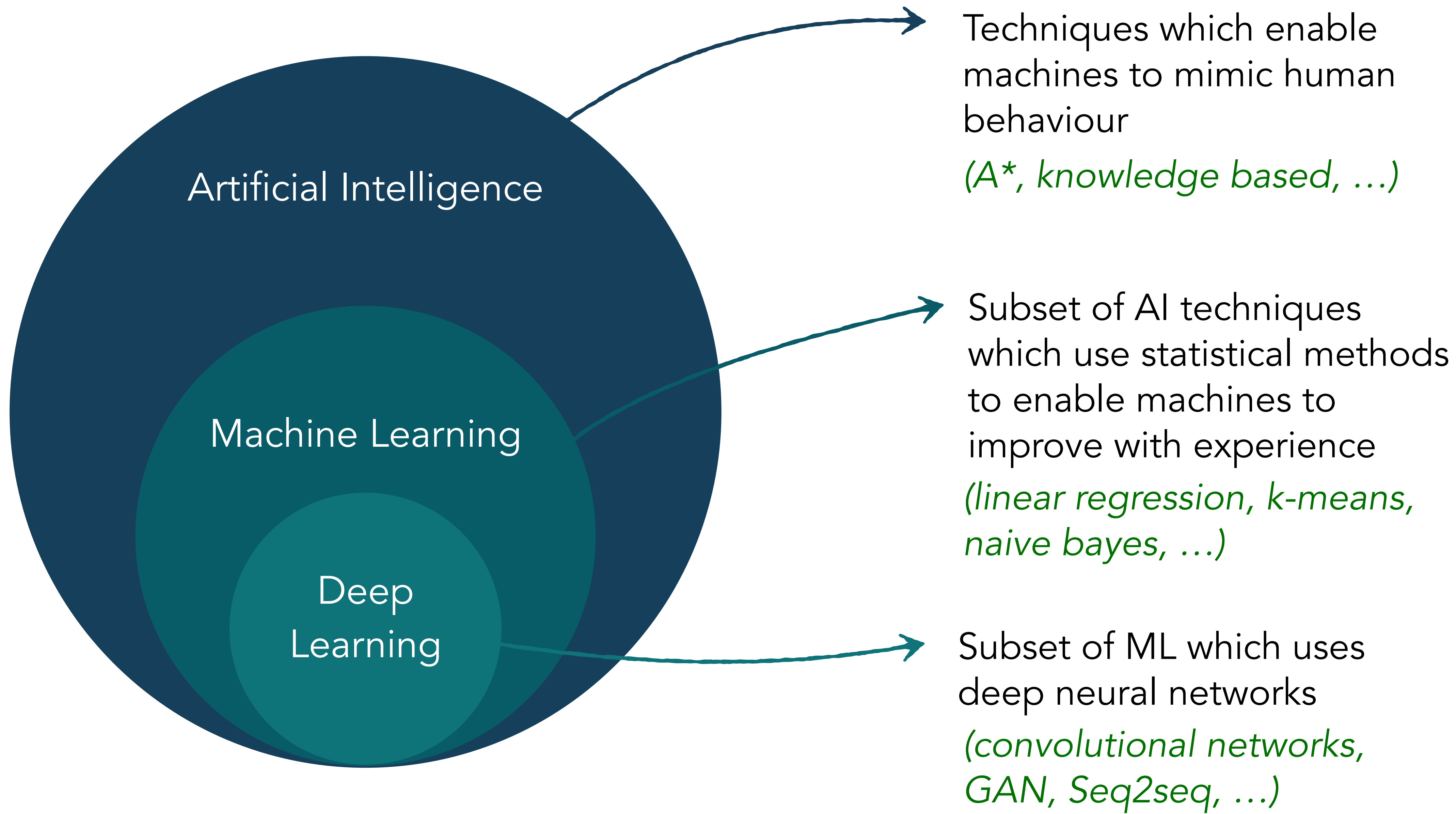


August 2022

We introduce `pharo-ai` v0.8

a modular library for shallow machine
learning in Pharo

 github.com/pharo-ai



Machine Learning



Supervised

- Labeled data
- First learn from examples, then apply to new data

(classification, regression,...)

Unsupervised

- No labeled data
- Extract patterns from the data

(clustering, anomaly detection,...)

Why do we need a ML library in Pharo

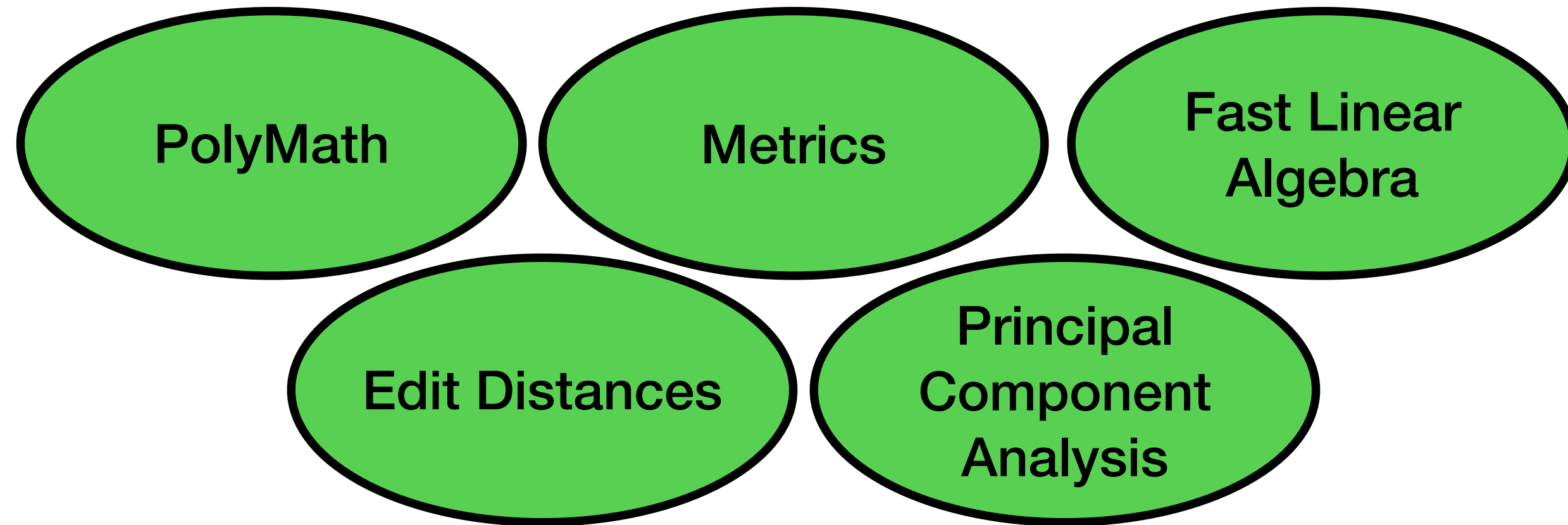
- We want to provide tools for the Pharo community people interested in doing ML and AI.
- We would like to contribute to the work that is currently being developed by different people (Univ. Chile, Object Profile-Chile, PolyMathOrg, Semantics-Bolivia, CIRAD-France).

How do we position ourselves

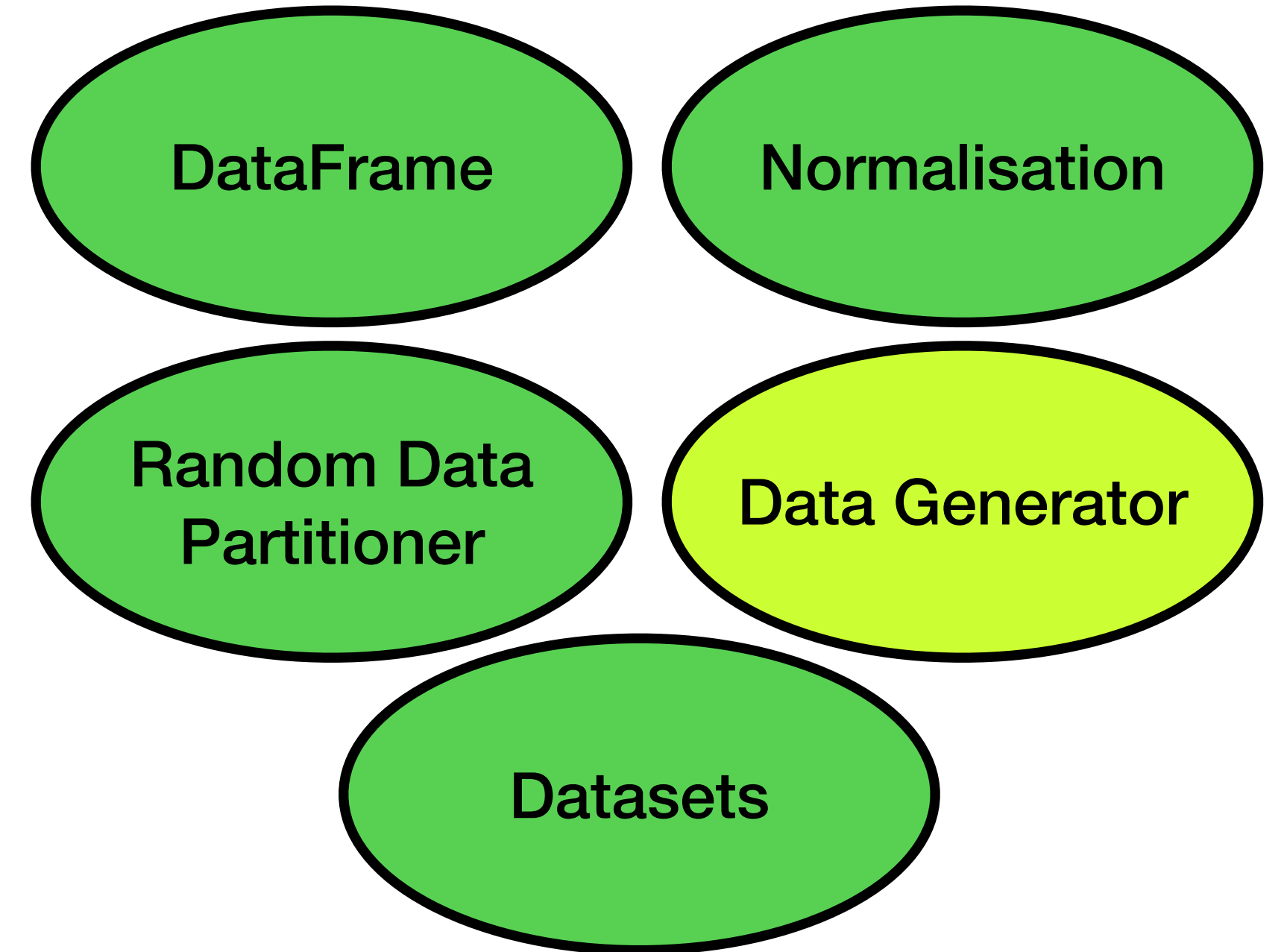
	Python	R	Pharo
Data Analysis & Manipulation	pandas	data.frame, dplyr	DataFrame
Algebra & Statistics	numpy, scipy	MASS, SparseM	PolyMath
Shallow Learning	scikit-learn	caret, ml3	pharo-ai
Deep Learning	TensorFlow, Keras	TensorFlow, Keras	TensorFlow, Keras
Visualisation	matplotlib	ggplot	Roassal

Algorithms and libraries available in
pharo-ai ecosystem

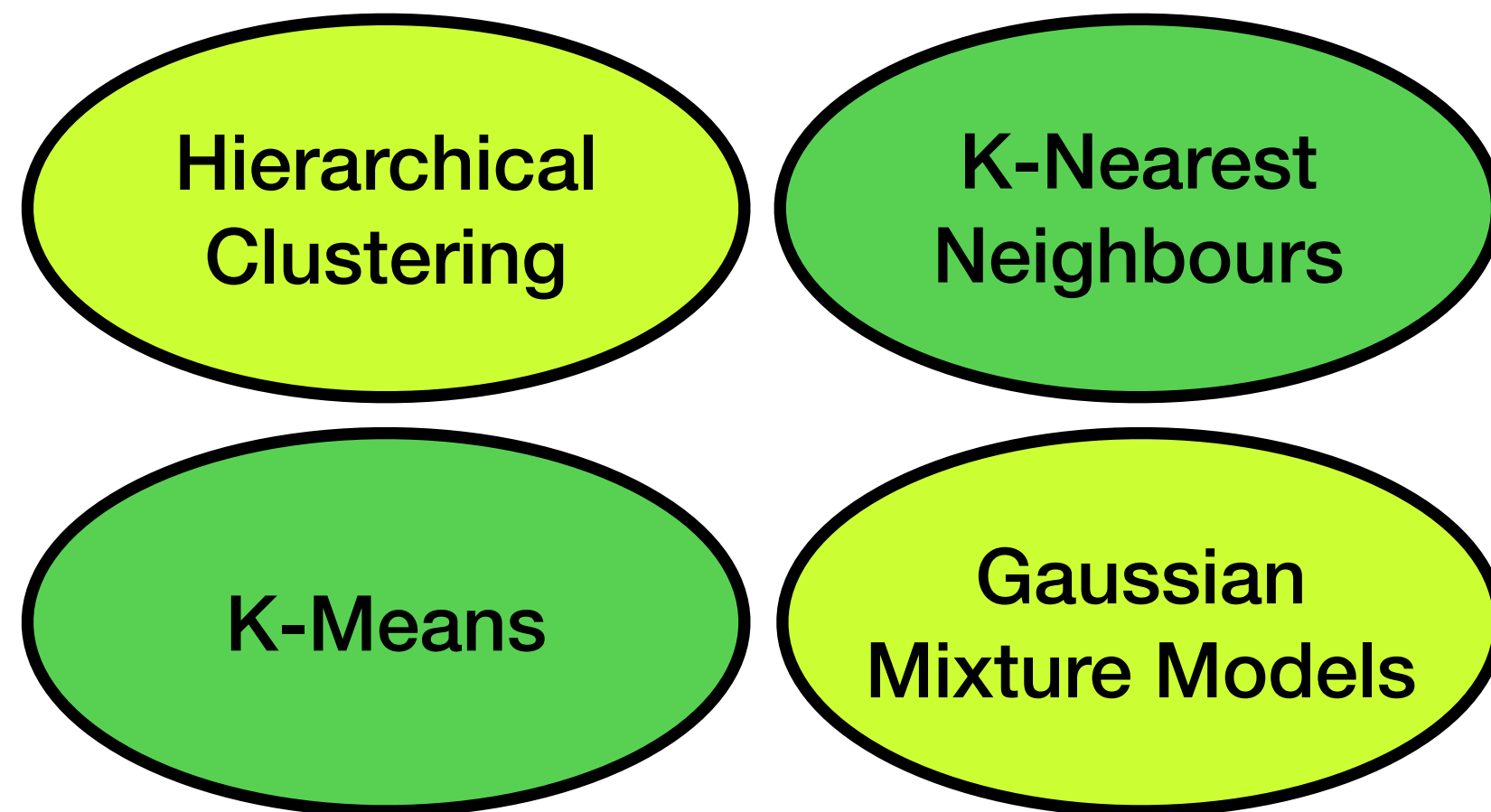
Scientific Computation



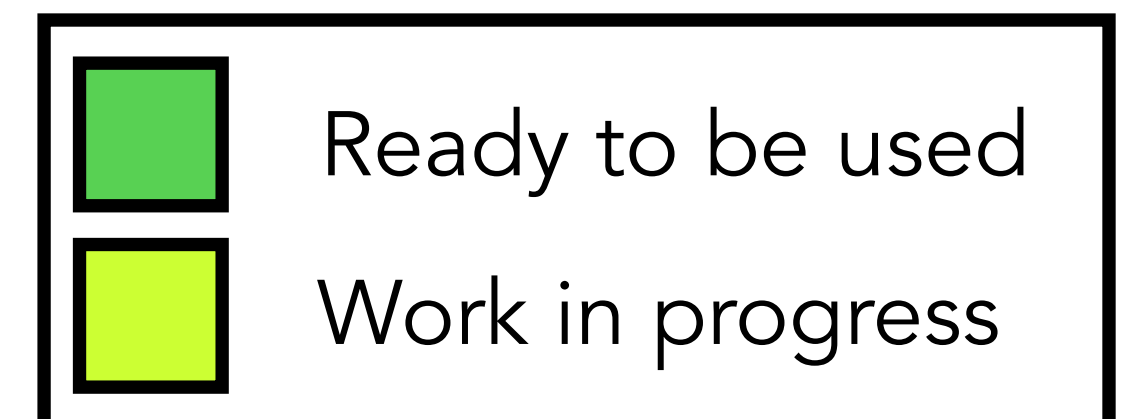
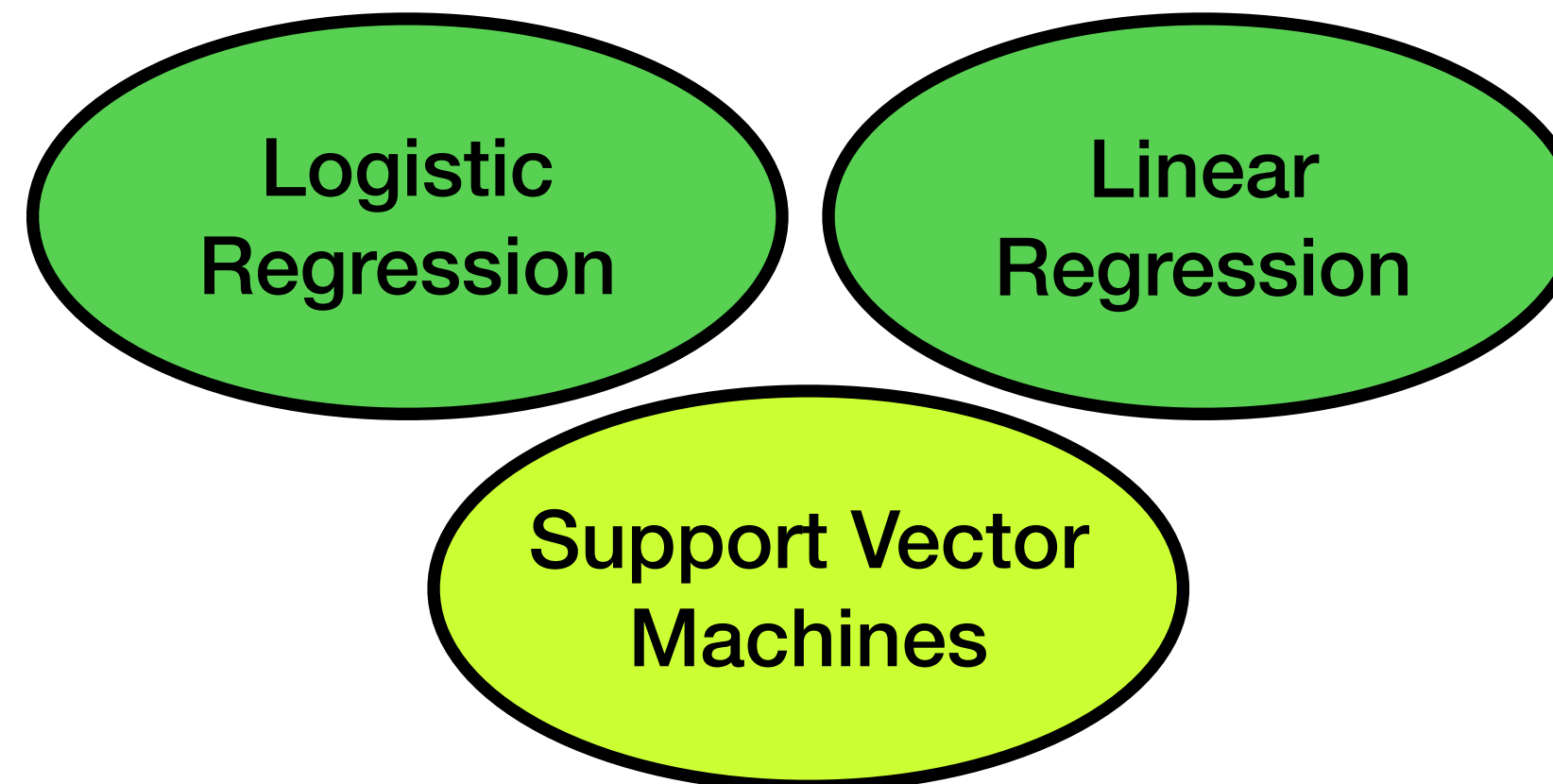
Data manipulation



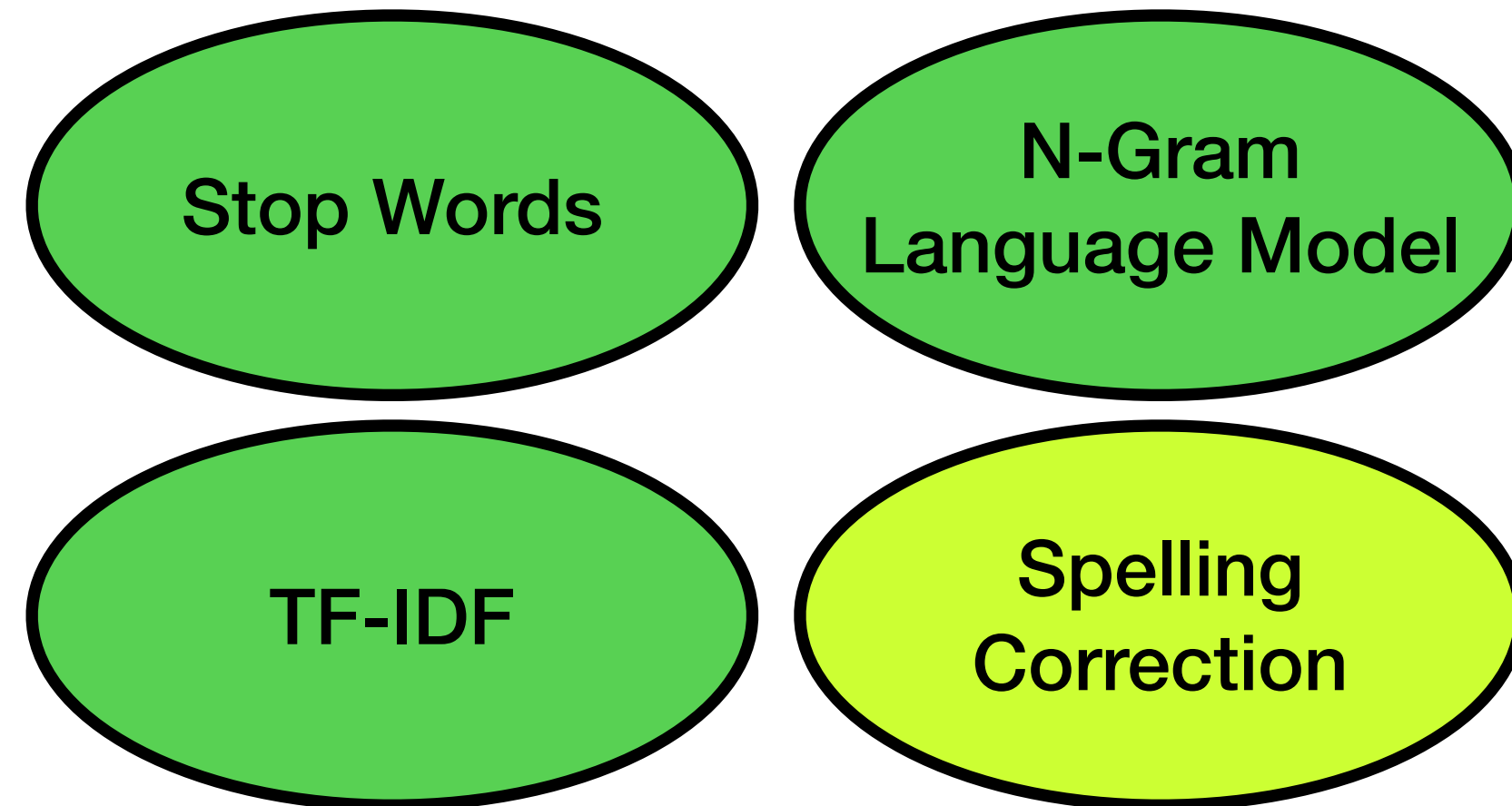
Clustering



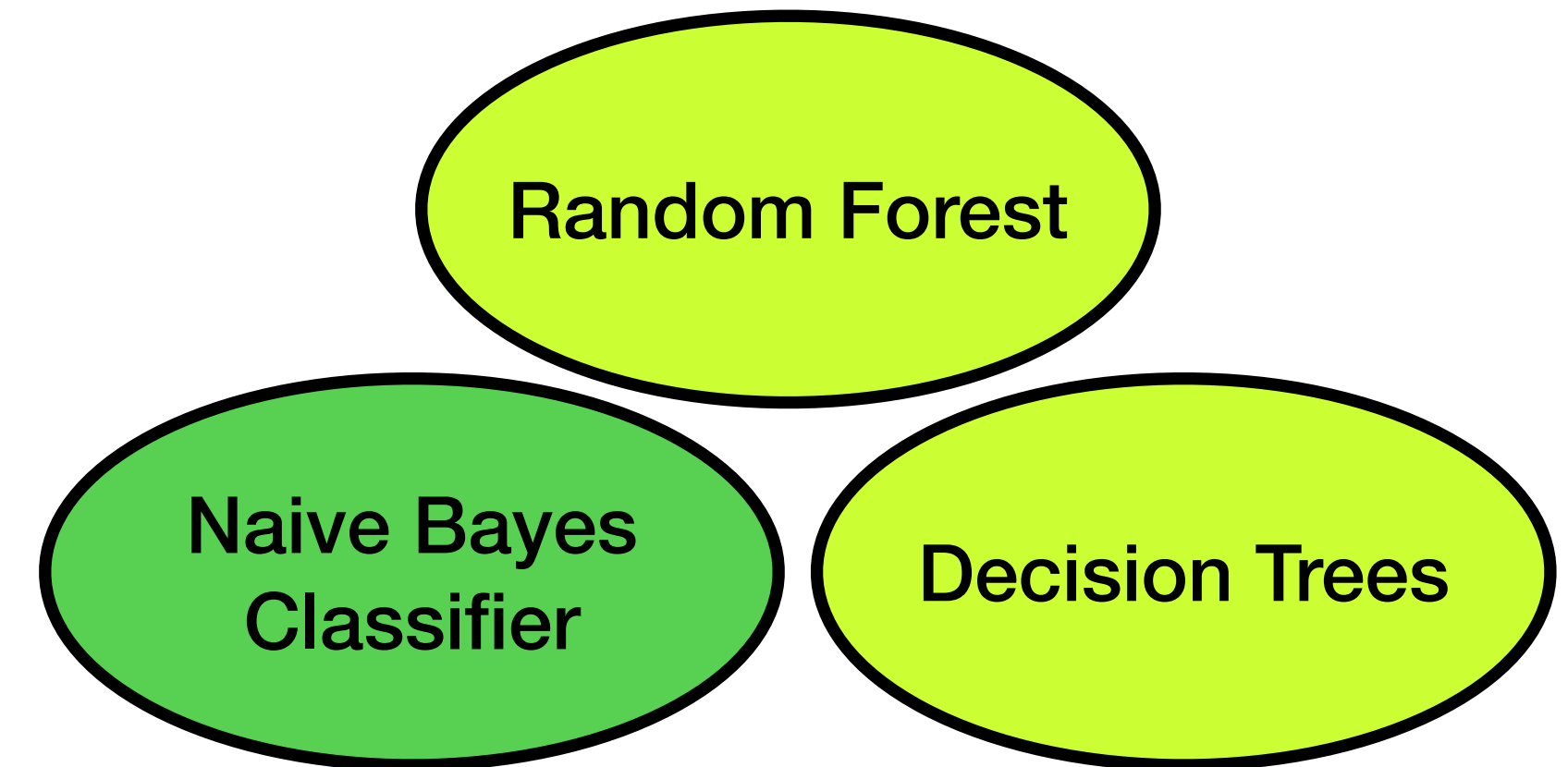
Regression



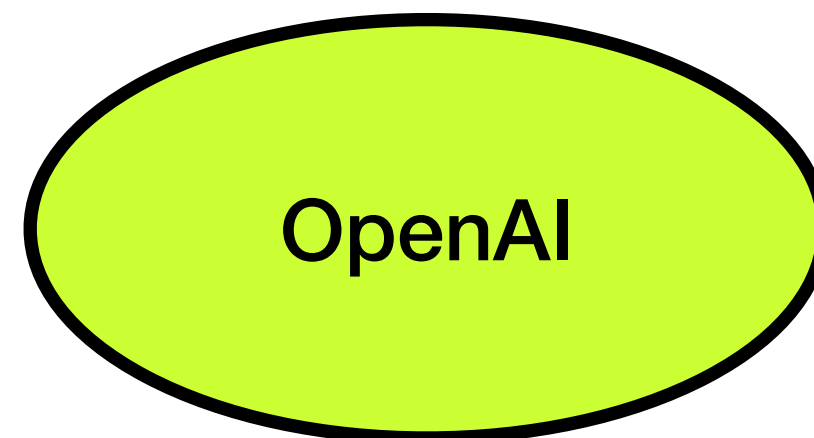
Natural Language Processing



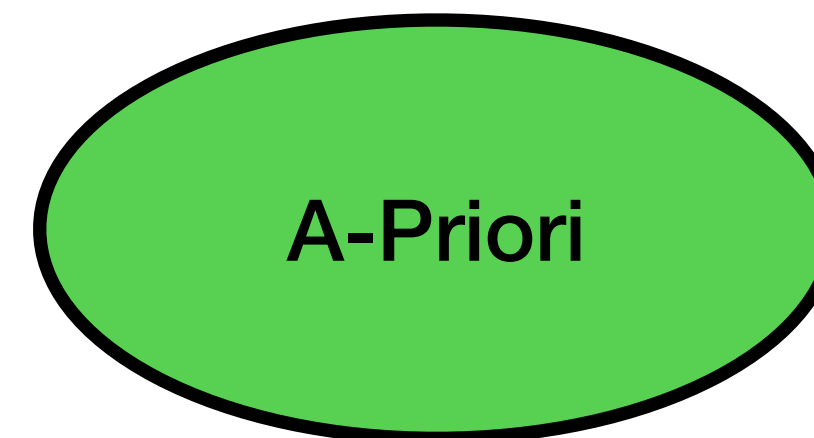
Classification



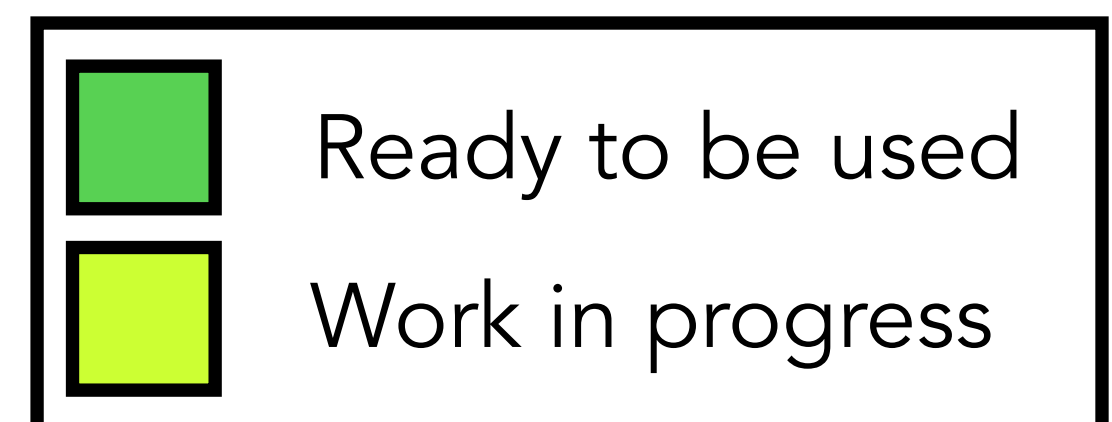
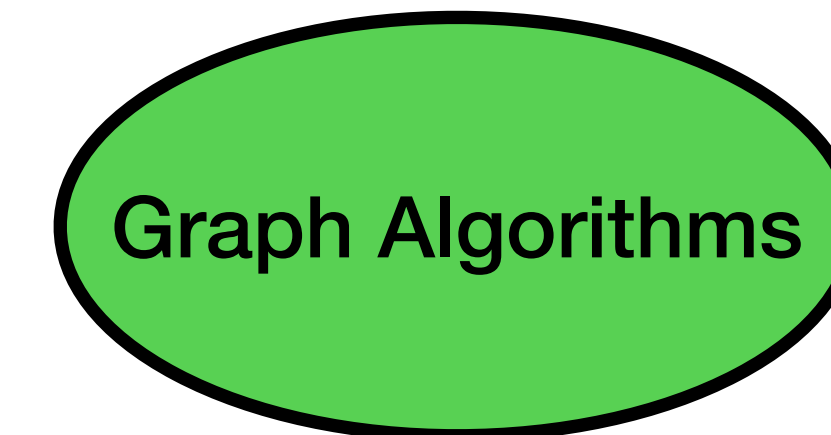
OpenAI Binding



Data Mining



Graph Algorithms



Scientific Computation

PolyMath

Metrics

Fast
Linear Algebra

Edit
Distances

Principal
Component

Classifica

Random
Forest

Naive
Bayes

Decision
Trees

Data

DataFrame

Normalisati
on

Random Data

Data
Generator

Datasets

Graph

Graph
Algorithms

Data Mining

A-Priori

Clustering

Hierarchical
Clustering

K-
Nearest

K-Means

Gaussian
Mixture Models

Natural Language

Stop Words

N-
Gram

TF-IDF

Spelling
Correction

Regression

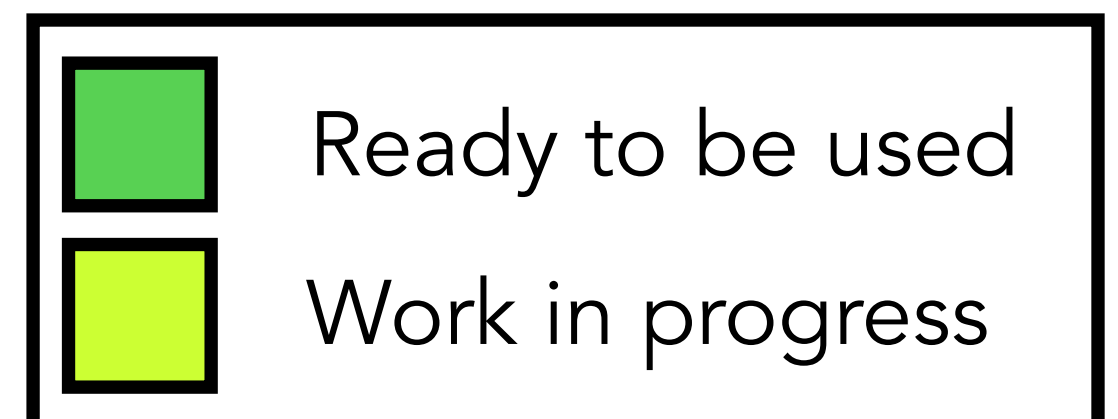
Logistic
Regression

Linear
Regression

Support
Vector Machines

OpenAI

OpenAI



Inspector Extensions for DataFrame

Inspector on a DataFrame [150 items] (a DataFrame('sepal length (cm)'->5.1 'sepal width (cm)'->3.5 'petal length (cm)'->1.4 'petal width (cm)'->0.2 'species'-'setosa') a DataFrame('sepal length (cm)'->4.9 'sepal width (cm)'->3.0 'petal length (cm)'->1.4 'petal width (cm)'->0.2 'species'-'setosa') a DataFrame [150 items] (a ...)

DataFrame Data Description Visualizations Raw Breakpoints Meta

sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7000000000000002	0.4	setosa
4.6	3.4	1.4	0.30000000000000004	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa

Statistic	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
1st Quartile	5.10	2.80	1.60	0.30	NaN
3rd Quartile	6.40	3.30	5.10	1.80	NaN
Median	5.80	3.00	4.35	1.30	NaN
Minimum	4.30	2.00	1.00	0.10	NaN
Maximum	7.90	4.40	6.90	2.50	NaN
Variance	0.69	0.19	3.12	0.58	NaN
Standard deviation	0.83	0.44	1.77	0.76	NaN
Mode	5.00	3.00	1.40	0.20	NaN
Average	5.84	3.06	3.76	1.20	NaN

Property	Value
Dimensions	(150@5)
Has categorical	true
Has nil	false

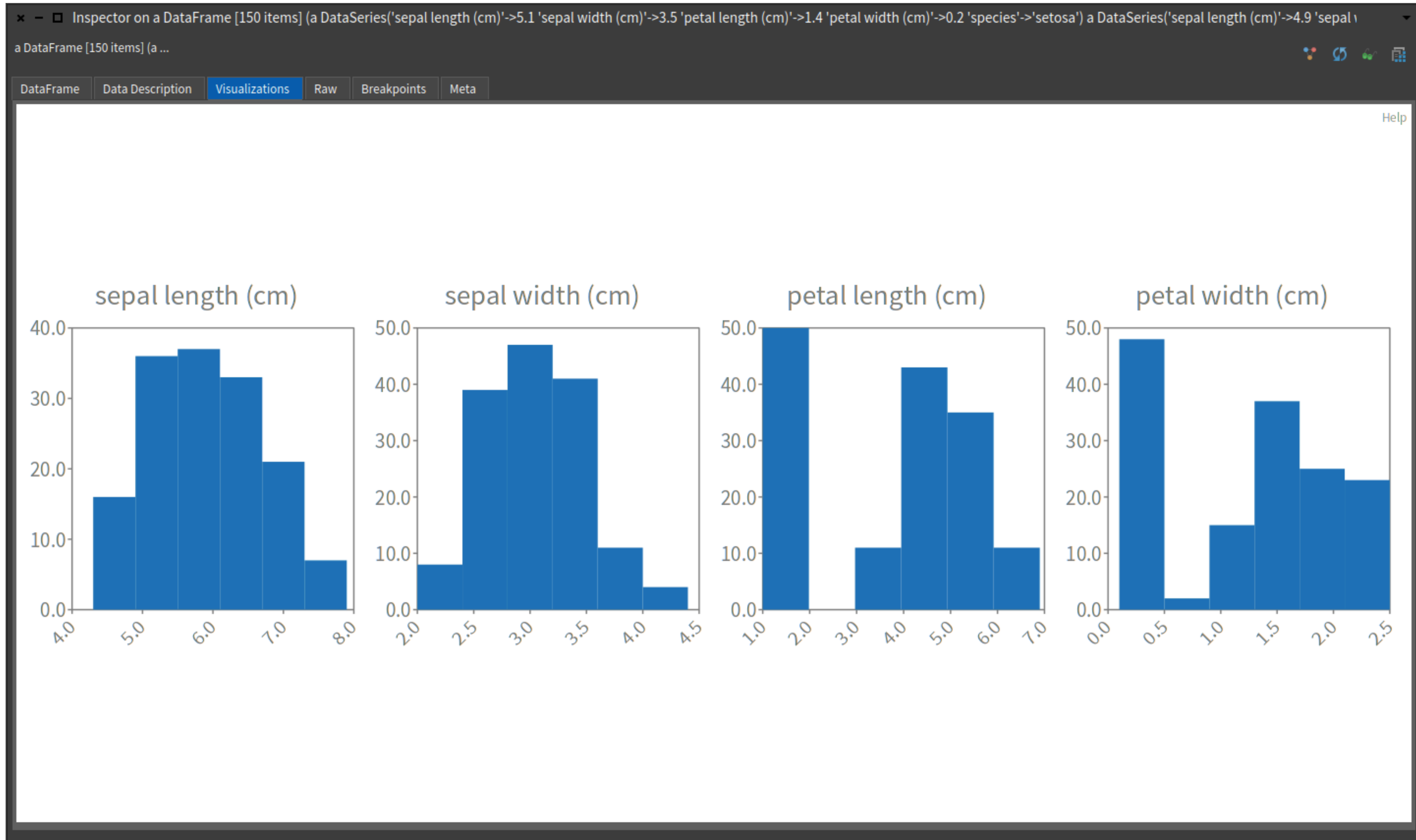
Do it

```

1 self.columnNames "an OrderedCollection('sepal length (cm)' 'sepal
width (cm)' 'petal length (cm)' 'petal width (cm)' 'species')".
2 (self.column: 'species') values asSet size "3"

```

Inspector Extensions for DataFrame



Other Machine Learning Projects

Awesome Machine Learning in Pharo

This is the list of machine learning projects written in or related to Pharo as well as books, booklets, papers, or tutorials on this topic. If you want to add an entry to this list - feel free to make a pull request or create an issue with a link to the project and we will add it ourselves. Entries are grouped into categories and sorted by alphabet.

Contents

- [Mathematics](#)
- [Linear Models](#)
- [Neural Networks](#)
- [Deep Learning](#)
- [Neuroevolution](#)
- [Generative Models](#)
- [Natural Language Processing](#)
- [Data Mining](#)
- [Data Structures](#)
- [Data Visualization](#)
- [Interactive Notebooks](#)
- [Data Collection](#)
- [Applications](#)
- [Books & Booklets](#)
- [Papers](#)

pharo-ai wiki

Pharo-AI Wiki

This is the Pharo-AI Wiki. The goal of this wiki is to provide documentation and tutorials to help people start using our Pharo AI/Machine-Learning libraries.

- [Getting Started page.](#)
- [Contributing guide](#)

If you want to see other Machine Learning projects in Pharo, please see: <https://github.com/pharo-ai/awesome-pharo-ml>

Keep in mind that the wiki and pharo-ai is right now under construction version so not all the algorithms will be documented or with all the functionalities that we would like to have. Nevertheless, all the things that are documented here had been revised and are working.

Contents

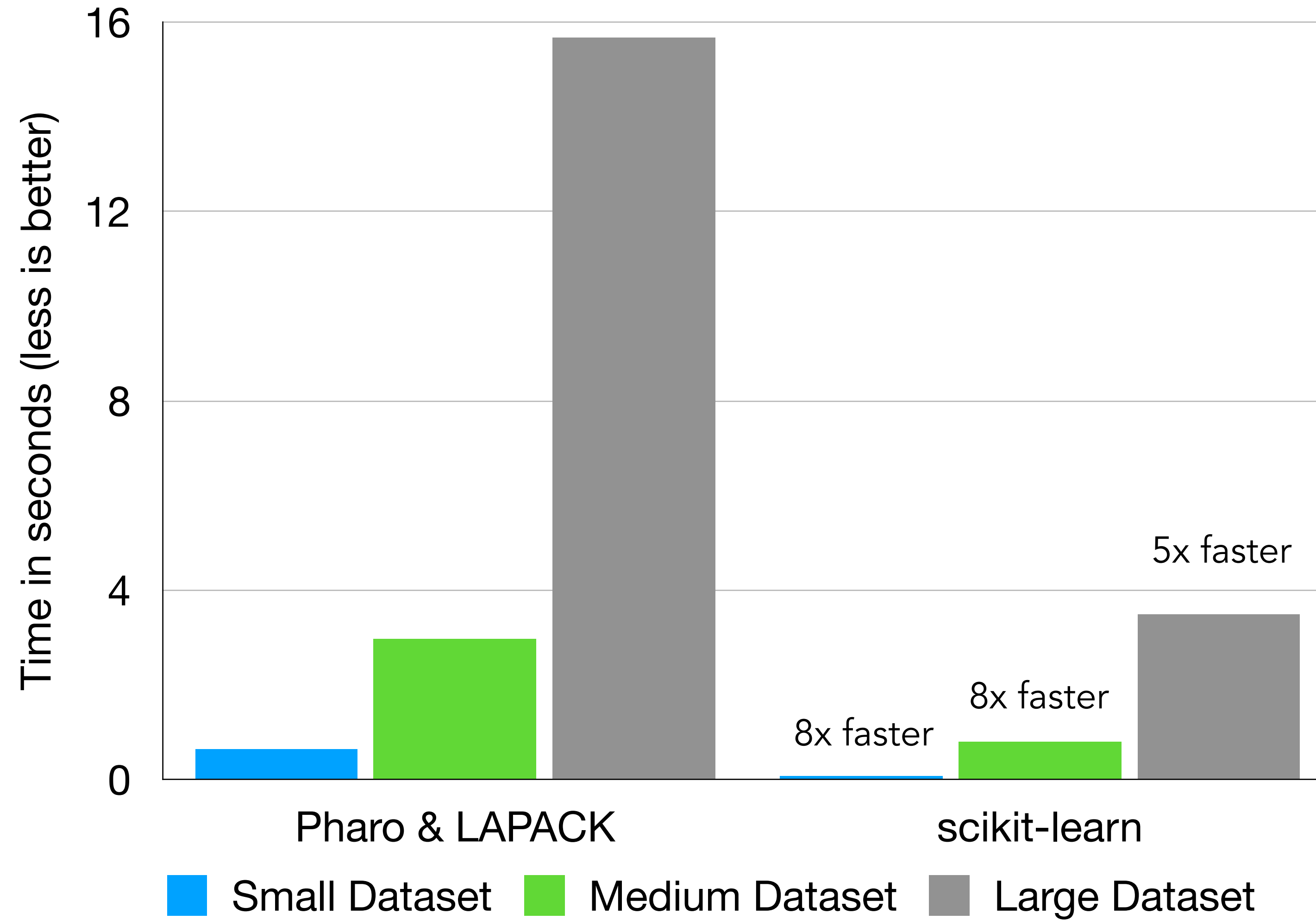
- [Tutorials](#)
 - [Linear Regression](#)
 - [Logistic Regression](#)
 - [Clustering](#)
 - [Edit Distances](#)
- [Machine Learning](#)
 - [Regression](#)
 - [Classification](#)
 - [Clustering](#)
 - [Using metrics](#)
- [Linear Algebra](#)
- [Data Preprocessing](#)
- [Data Mining](#)

Linear Regression benchmarks

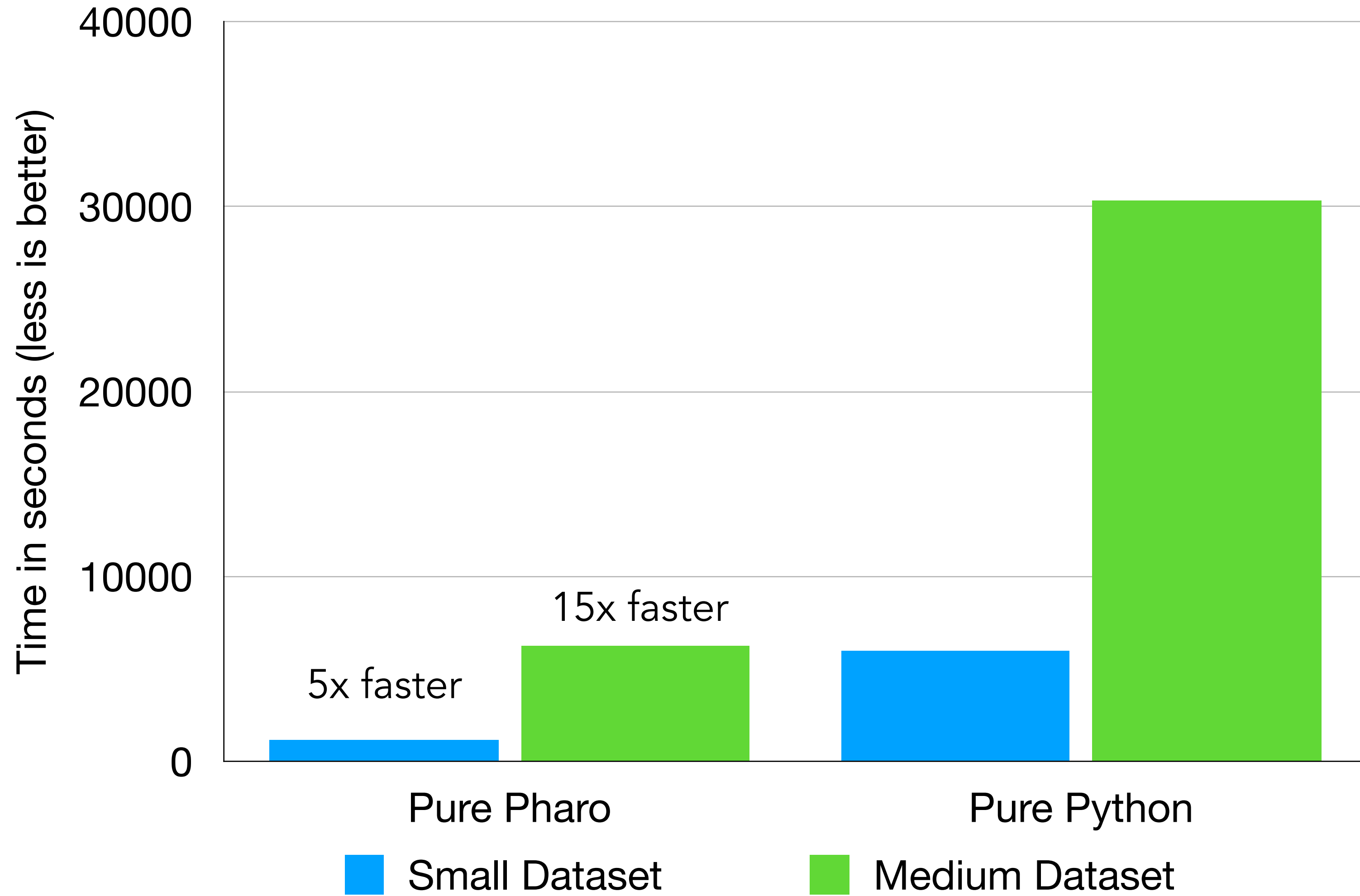
Dataset size

Name	Columns	Rows	Size
Small	200,000	20	82 Mb
Medium	1,000,000	20	411 Mb
Large	5,000,000	20	2,06 Gb

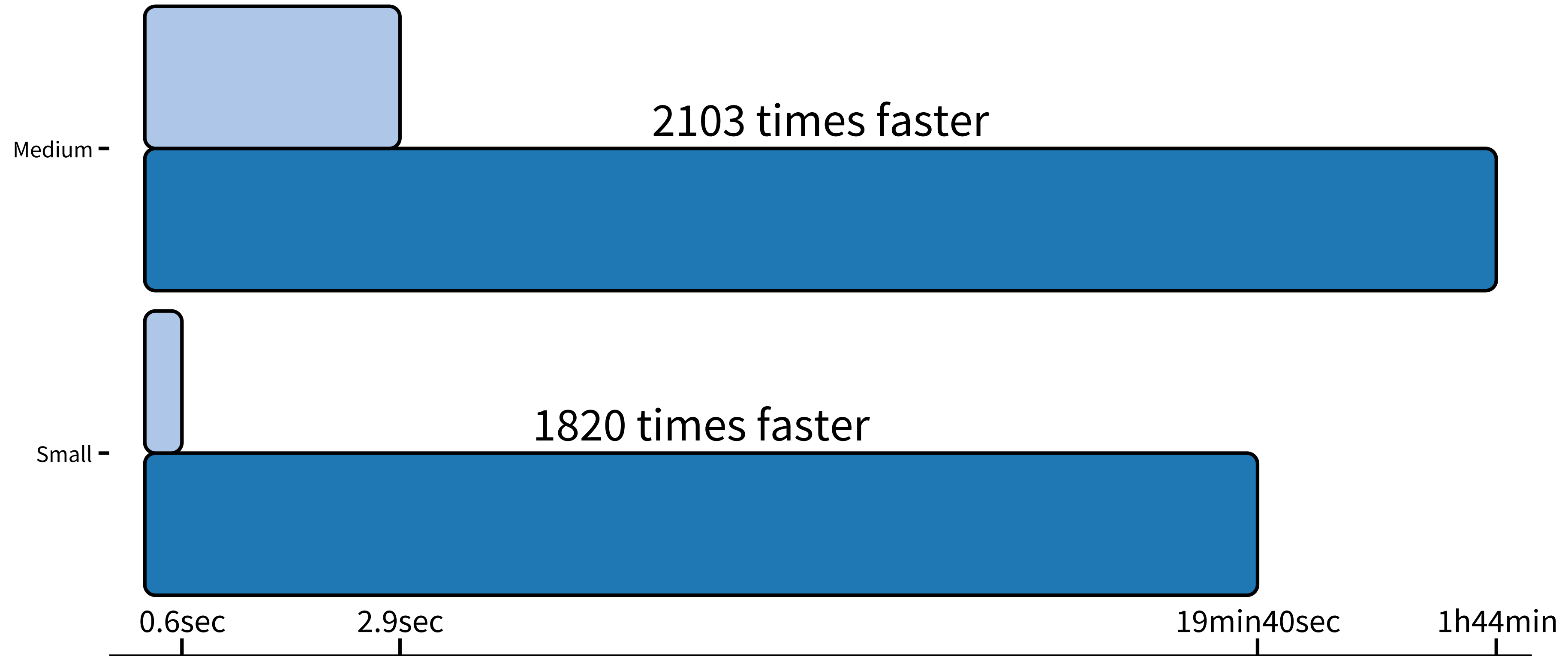
Pharo & LAPACK vs Scikit-learn



Pure Pharo vs Python



Pure Pharo vs Pharo & LAPACK



Time in seconds (less is better)



Visit Us ! Play, Use, and Contribute

 **Start here**

pharo-ai wiki: <https://github.com/pharo-ai/wiki>

