Fundaments of Machine learning for and with engineering applications

Enrico Riccardi¹

Department of Mathematics and Physics, University of Stavanger (UiS). 1

Sep 3, 2025



© 2025, Enrico Riccardi. Released under CC Attribution 4.0 license

1 Statistics, Machine learning or Artificial intelligence?

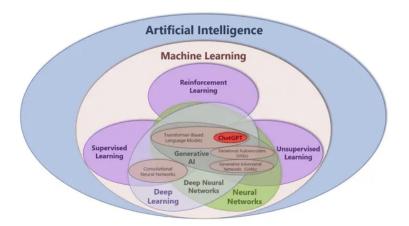
2 Machine Learning intro

Coding quality

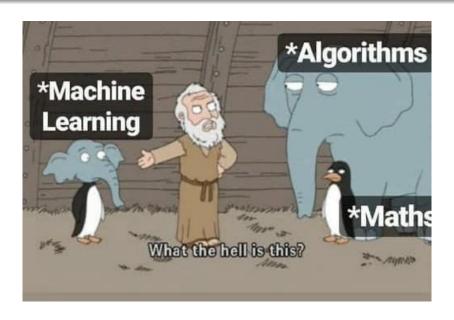
4 Version contro

Statistics, Machine learning or Artificial intelligence?

What is the main difference between the three fields?



How Machine Learning Started?



- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments. [Wikipedia]

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments. [Wikipedia]

Definitions:

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]
- Machine learning is a subfield of artificial intelligence that uses algorithms
 trained on data sets to create models that enable machines to perform tasks
 that would otherwise only be possible for humans, such as categorizing images,
 analyzing data, or predicting price fluctuations. [Coursera]

Definitions

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]
- Machine learning is a subfield of artificial intelligence that uses algorithms
 trained on data sets to create models that enable machines to perform tasks
 that would otherwise only be possible for humans, such as categorizing images,
 analyzing data, or predicting price fluctuations. [Coursera]

Definitions

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]
- Machine learning is a subfield of artificial intelligence that uses algorithms
 trained on data sets to create models that enable machines to perform tasks
 that would otherwise only be possible for humans, such as categorizing images,
 analyzing data, or predicting price fluctuations. [Coursera]

Definitions

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]
- Machine learning is a subfield of artificial intelligence that uses algorithms trained on data sets to create models that enable machines to perform tasks that would otherwise only be possible for humans, such as categorizing images, analyzing data, or predicting price fluctuations. [Coursera]

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning

Loss function

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning

Loss function

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- Artificial intelligence (Al) is the theory and development of computer systems capable of performing tasks that historically required human intelligence, such as recognizing speech, making decisions, and identifying patterns [Coursera]
- It is the science and engineering of making intelligent machines, especially
 intelligent computer programs. It is related to the similar task of using
 computers to understand human intelligence, but AI does not have to confine
 itself to methods that are biologically observable. [IBM]

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- Artificial intelligence (AI) is the theory and development of computer systems capable of performing tasks that historically required human intelligence, such as recognizing speech, making decisions, and identifying patterns [Coursera]
- It is the science and engineering of making intelligent machines, especially
 intelligent computer programs. It is related to the similar task of using
 computers to understand human intelligence, but AI does not have to confine
 itself to methods that are biologically observable. [IBM]

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- Artificial intelligence (AI) is the theory and development of computer systems capable of performing tasks that historically required human intelligence, such as recognizing speech, making decisions, and identifying patterns [Coursera]
- It is the science and engineering of making intelligent machines, especially
 intelligent computer programs. It is related to the similar task of using
 computers to understand human intelligence, but Al does not have to confine
 itself to methods that are biologically observable. [IBM]

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- Artificial intelligence (AI) is the theory and development of computer systems capable of performing tasks that historically required human intelligence, such as recognizing speech, making decisions, and identifying patterns [Coursera]
- It is the science and engineering of making intelligent machines, especially
 intelligent computer programs. It is related to the similar task of using
 computers to understand human intelligence, but AI does not have to confine
 itself to methods that are biologically observable. [IBM]

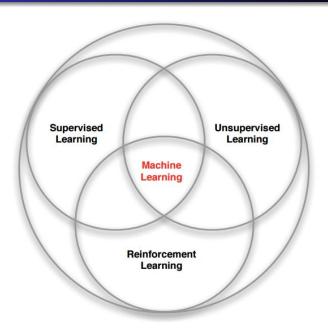
1 Statistics, Machine learning or Artificial intelligence?

Machine Learning intro

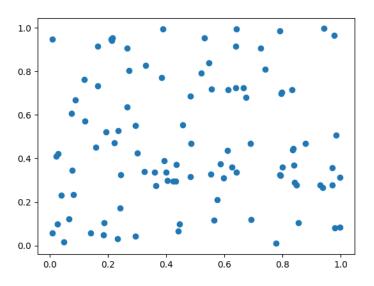
Coding quality

4 Version contro

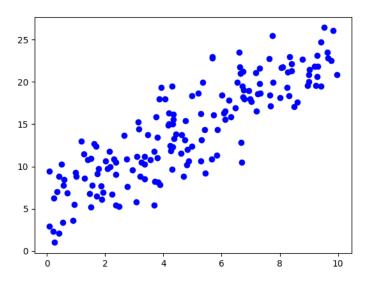
Families of Machine learning



What can we do with that?



What about in this case?



Python Source code 1

```
import numpy as np
import matplotlib.pyplot as plt

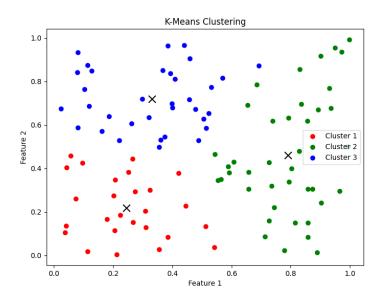
# Generate some sample data
data = np.random.rand(100, 2) # 100 data points with 2 features

plt.scatter(data[:, 0], data[:, 1])
plt.show()
```

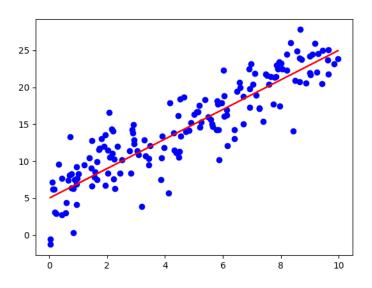
Python Source code 2

```
import numpy as np
import matplotlib.pyplot as plt
def generate_linear_data(n_random_points, noise=16):
    x = np.random.rand(n_random_points) * 10
    # Make 'perfect' data
    true_slope, true_intercept = 2, 5
    y = true_slope * x + true_intercept
    # Add noise
    y += np.random.randn(n_random_points)*noise
    return x, y, true_slope, true_intercept
# Use the function to generate data
x, y, true_slope, true_intercept = generate_linear_data(
        n_random_points=166,
        noise=3)
# Plot all
plt.scatter(x, y, color='blue', label='Data Points')
plt.show()
```

Unsupervised learning



Supervised learning



The data decides

This is why we focus so much on the data type.

The data properties dictate what statistical model can be adopted.

An statistical model has leverages our understanding of the data structure to improve its **predictions** (inference).

The numerical recipe that we used to generate the data is defined the truth

Psychology or data science?

Most Machine learning tools are aimed to find the truth. In most cases, we are happy to not find lies.

The data decides

This is why we focus so much on the data type.

The data properties dictate what statistical model can be adopted.

An statistical model has leverages our understanding of the data structure to improve its **predictions** (inference).

The numerical recipe that we used to generate the data is defined the truth

Psychology or data science?

Most Machine learning tools are aimed to find the truth. In most cases, we are happy to not find lies.

Unsupervised learning

Unsupervised learning, a term that resonates with the autonomy of machine intelligence, operates on the principle of identifying patterns and structures in datasets without labelled responses.

This branch of machine learning is distinguished by its lack of explicit guidance, where algorithms are tasked with uncovering hidden structures from unlabeled data.

The most common clustering strategies are

- filtering
- clustering
- dimensionality reduction
- association learning

Unsupervised learning

Unsupervised learning, a term that resonates with the autonomy of machine intelligence, operates on the principle of identifying patterns and structures in datasets without labelled responses.

This branch of machine learning is distinguished by its lack of explicit guidance, where algorithms are tasked with uncovering hidden structures from unlabeled data.

The most common clustering strategies are :

- filtering
- clustering
- dimensionality reduction
- association learning

Application of unsupervised learning

It is a bit of a holy grail: a computer that finds patterns without guidance. (Yes, it doesn't work, most of the time)

Still, it has been shown efficient for

- Computer vision
- Anomaly detection
- Exploratory data analysis

Main challenge

The right result is quite undefined, Uncertain goa

We will demonstrate it with a famous problem.

Application of unsupervised learning

It is a bit of a holy grail: a computer that finds patterns without guidance. (Yes, it doesn't work, most of the time)

Still, it has been shown efficient for:

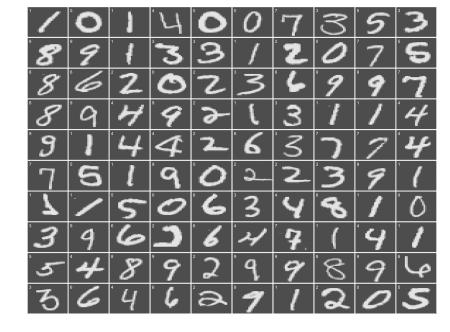
- Computer vision
- Anomaly detection
- Exploratory data analysis

Main challenge

The right result is quite undefined, Uncertain goal.

We will demonstrate it with a famous problem.

Uncertain goal



Weak supervised learning

 \boldsymbol{A} less popular type of machine learning problem is when labels are assigned to groups of instances.

The group of instances is called bag.

The question is, what is the level of a previously unforeseen bag?

This data structure and question type request a hybrid treatment between supervised and supervised learning.

Multiple instance learing

Multiple instances are needed to learn (quite clear name)

Weak supervised learning

 \boldsymbol{A} less popular type of machine learning problem is when labels are assigned to groups of instances.

The group of instances is called bag.

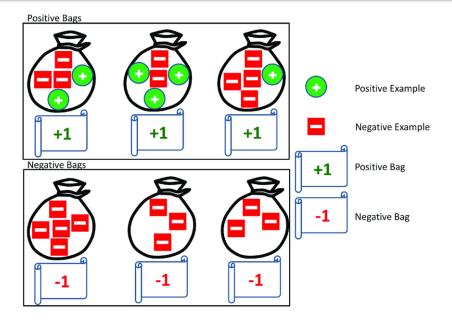
The question is, what is the level of a previously unforeseen bag?

This data structure and question type request a hybrid treatment between supervised and supervised learning.

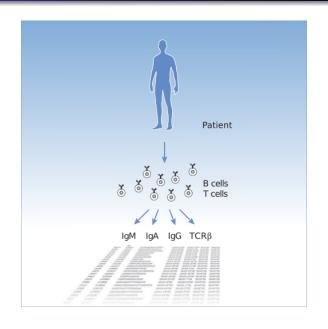
Multiple instance learing

Multiple instances are needed to learn (quite clear name)

Weak Supervised learning



Weak Supervised learning



Reinforcement learning

Finally, there is a further approach.

Reinforcement learning (RL)

It aims to train an intelligent agent to take actions in a dynamic environment in order to maximise the cumulative reward.

It learns from outcomes and decides which action to take next. After each action, the algorithm receives feedback that helps it determine whether the choice it made was correct, neutral or incorrect.

It is a self-teaching system that essentially learns by trial and error

It is a dependable tool for automated decision making

Reinforcement learning

Finally, there is a further approach.

Reinforcement learning (RL)

It aims to train an intelligent agent to take actions in a dynamic environment in order to maximise the cumulative reward.

It learns from outcomes and decides which action to take next. After each action, the algorithm receives feedback that helps it determine whether the choice it made was correct, neutral or incorrect.

It is a self-teaching system that essentially learns by trial and error.

It is a dependable tool for automated decision making.

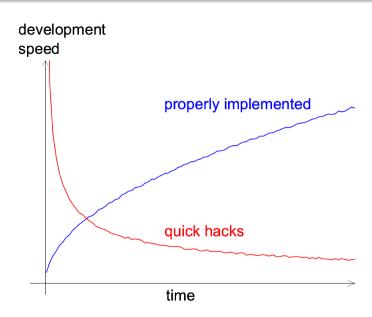
Statistics, Machine learning or Artificial intelligence?

2 Machine Learning intro

3 Coding quality

4 Version contro

Flexible



- jupyter notebooks are mostly dedicated to learning (Markdown)
- ipython is for interactive coding (similar to R, Matlab, etc)
- python packages (.py) developing suites (debug possibilities and git integration)

- jupyter notebooks are mostly dedicated to learning (Markdown)
- ipython is for interactive coding (similar to R, Matlab, etc)
- python packages (.py) developing suites (debug possibilities and git integration)

- jupyter notebooks are mostly dedicated to learning (Markdown)
- ipython is for interactive coding (similar to R, Matlab, etc)
- python packages (.py) developing suites (debug possibilities and git integration)

- jupyter notebooks are mostly dedicated to learning (Markdown)
- ipython is for interactive coding (similar to R, Matlab, etc)
- python packages (.py) developing suites (debug possibilities and git integration)

Introducing code standards

When developing code, there are **guidelines** and best practices aimed at improving the quality, readability, and maintainability of a code.

There are different levels of coding quality, mostly depending on the code intended usage (and developer skills).

- Private codes can be whatever (Cpt. Obvious)
- Public packages shall use a 'Golden code standards' such to be used and eventually supported by communities.

Introducing code standards

When developing code, there are **guidelines** and best practices aimed at improving the quality, readability, and maintainability of a code.

There are different levels of coding quality, mostly depending on the code intended usage (and developer skills).

- Private codes can be whatever (Cpt. Obvious)
- Public packages shall use a 'Golden code standards' such to be used and eventually supported by communities.

- Readability and Clarity: A good code shall be possible to read as when reading a book
- Structure and object oriented: A code shall be composed by objects, each of them connected in the less redundant way possible.
- Consistency and Style: Variable naming, function naming and classes naming has to be consistent.
- Documentation: Each file, each function and each class shall contain the relative description of its aim and its usage
- Maintainability. Code dependencies have to be stated and consistently defined and updated, such that a suitable environment can be developed at any point in time

- Readability and Clarity: A good code shall be possible to read as when reading a book
- Structure and object oriented: A code shall be composed by objects, each of them connected in the less redundant way possible.
- Consistency and Style: Variable naming, function naming and classes namin has to be consistent.
- Occumentation: Each file, each function and each class shall contain the relative description of its aim and its usage
- Maintainability: Code dependencies have to be stated and consistently defined and updated, such that a suitable environment can be developed at any point in time.

- Readability and Clarity: A good code shall be possible to read as when reading a book
- 2 Structure and object oriented: A code shall be composed by objects, each of them connected in the less redundant way possible.
- Consistency and Style: Variable naming, function naming and classes naming has to be consistent.
- Occumentation: Each file, each function and each class shall contain the relative description of its aim and its usage
- Maintainability: Code dependencies have to be stated and consistently defined and updated, such that a suitable environment can be developed at any point in time.

- Readability and Clarity: A good code shall be possible to read as when reading a book
- 2 Structure and object oriented: A code shall be composed by objects, each of them connected in the less redundant way possible.
- Consistency and Style: Variable naming, function naming and classes naming has to be consistent.
- Documentation: Each file, each function and each class shall contain the relative description of its aim and its usage
- Maintainability: Code dependencies have to be stated and consistently defined and updated, such that a suitable environment can be developed at any point in time.

- Readability and Clarity: A good code shall be possible to read as when reading a book
- Structure and object oriented: A code shall be composed by objects, each of them connected in the less redundant way possible.
- Consistency and Style: Variable naming, function naming and classes naming has to be consistent.
- Occumentation: Each file, each function and each class shall contain the relative description of its aim and its usage
- Maintainability: Code dependencies have to be stated and consistently defined and updated, such that a suitable environment can be developed at any point in time.

- Readability and Clarity: A good code shall be possible to read as when reading a book
- Structure and object oriented: A code shall be composed by objects, each of them connected in the less redundant way possible.
- Consistency and Style: Variable naming, function naming and classes naming has to be consistent.
- Occumentation: Each file, each function and each class shall contain the relative description of its aim and its usage
- Maintainability: Code dependencies have to be stated and consistently defined and updated, such that a suitable environment can be developed at any point in time.

- 1 Testing: Unit testing shall cover the majority of the code
- 2 Error Handling: Each error shall be captured and properly identified
- Examples and benchmarks: Users shall be able to execute minimal examples of the code for computational checks.
- Performance Optimization: Libraries shall be able to use the available computational power in the machine (e.g. GPU-CUDA)

- Testing: Unit testing shall cover the majority of the code
- 2 Error Handling: Each error shall be captured and properly identified
- ② Examples and benchmarks: Users shall be able to execute minimal examples of the code for computational checks.
- Performance Optimization: Libraries shall be able to use the available computational power in the machine (e.g. GPU-CUDA)

- Testing: Unit testing shall cover the majority of the code
- 2 Error Handling: Each error shall be captured and properly identified.
- Sexamples and benchmarks: Users shall be able to execute minimal examples of the code for computational checks.
- Performance Optimization: Libraries shall be able to use the available computational power in the machine (e.g. GPU-CUDA)

- 1 Testing: Unit testing shall cover the majority of the code
- 2 Error Handling: Each error shall be captured and properly identified.
- Examples and benchmarks: Users shall be able to execute minimal examples of the code for computational checks.
- Performance Optimization: Libraries shall be able to use the available computational power in the machine (e.g. GPU-CUDA)

- 1 Testing: Unit testing shall cover the majority of the code
- 2 Error Handling: Each error shall be captured and properly identified.
- Examples and benchmarks: Users shall be able to execute minimal examples of the code for computational checks.
- Performance Optimization: Libraries shall be able to use the available computational power in the machine (e.g. GPU-CUDA)

1 Statistics, Machine learning or Artificial intelligence?

Machine Learning intro

Coding quality

Version control

"FINAL".doc







FINAL_rev.2.doc







FINAL_rev.6.COMMENTS.doc

FINAL_rev.8.comments5. CORRECTIONS.doc









FINAL_rev.18.comments7.

FINAL_rev.22.comments49. corrections 9. MORE. 30. doc corrections 10. #@\$ WHYDID ICOMETOGRADS CHOOL???? doc

Git

Git is a distributed version control system that tracks changes in any set of computer files, usually used for coordinating work among programmers who are collaboratively developing source code during software development. Its goals include speed, data integrity, and support for distributed, non-linear workflows (thousands of parallel branches running on different computers). [Wiki]

Let's try to be more accessible

Git is a computer program/tool to save and download files on a hosting server (e.g. GitHub and GitLab).

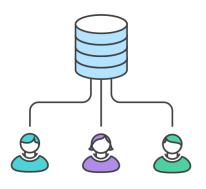
Git

Git is a distributed version control system that tracks changes in any set of computer files, usually used for coordinating work among programmers who are collaboratively developing source code during software development. Its goals include speed, data integrity, and support for distributed, non-linear workflows (thousands of parallel branches running on different computers). [Wiki]

Let's try to be more accessible.

Git is a computer program/tool to save and download files on a hosting server (e.g. GitHub and GitLab).

Centralized workflow



A distributed version control system

GIT

 Git facilitates users to track the various versions of files. It is not a necessary tool, but it can be very very helpful. Generally, the time spent to learn its syntax is well paid off

(do you remember to save some file like manuscript_draft_v4.02_final_definitive_forreal_lastcomments_editedbyER_submittedV Exactly! Imagine to do that for a repository of files...)

It permits to save and share the intermediate stages of a work in progress (which software is complete and always up to date?) in an accessible, consistent and structured way, allowing an effective version tracking. It allows retrieval of previous working versions, limiting the risk to overwrite useful files.

A distributed version control system

GIT

 Git facilitates users to track the various versions of files. It is not a necessary tool, but it can be very very helpful. Generally, the time spent to learn its syntax is well paid off

(do you remember to save some file like manuscript _draft _v4.02 _final _definitive _forreal _lastcomments _editedbyER _submittedVe Exactly! Imagine to do that for a repository of files...)

• It permits to save and share the intermediate stages of a work in progress (which software is complete and always up to date?) in an accessible, consistent and structured way, allowing an effective version tracking. It allows retrieval of previous working versions, limiting the risk to overwrite useful files.

A distributed version control system

GIT

 Git facilitates users to track the various versions of files. It is not a necessary tool, but it can be very very helpful. Generally, the time spent to learn its syntax is well paid off

(do you remember to save some file like manuscript _draft_v4.02 _final _definitive _forreal _lastcomments _editedbyER _submittedVeExactly! Imagine to do that for a repository of files...)

It permits to save and share the intermediate stages of a work in progress (which software is complete and always up to date?) in an accessible, consistent and structured way, allowing an effective version tracking. It allows retrieval of previous working versions, limiting the risk to overwrite useful files.

- Git helps to co-develop a code, test its functions and the compatibility of the various code sections.
- A long list of further possibilities became possible by git
- Different software integration on development platforms, based on git, will help you to develop and co-develop your code.
- The platform GitLab and GitHub have a large set of functionalities to further support code documentation and public releases.
- Files can be disclosed to the public, becoming a great integration of your CV showing what you are able to do in an open and accessible way.

- Git helps to co-develop a code, test its functions and the compatibility of the various code sections.
- A long list of further possibilities became possible by git.
- Different software integration on development platforms, based on git, will help you to develop and co-develop your code.
- The platform GitLab and GitHub have a large set of functionalities to furthe support code documentation and public releases.
- Files can be disclosed to the public, becoming a great integration of your CV showing what you are able to do in an open and accessible way.

- Git helps to co-develop a code, test its functions and the compatibility of the various code sections.
- A long list of further possibilities became possible by git.
- Different software integration on development platforms, based on git, will help you to develop and co-develop your code.
- The platform GitLab and GitHub have a large set of functionalities to further support code documentation and public releases.
- Files can be disclosed to the public, becoming a great integration of your CV, showing what you are able to do in an open and accessible way.

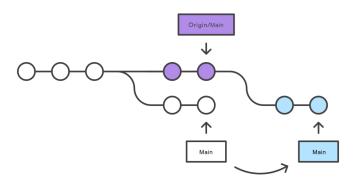
- Git helps to co-develop a code, test its functions and the compatibility of the various code sections.
- A long list of further possibilities became possible by git.
- Different software integration on development platforms, based on git, will help you to develop and co-develop your code.
- The platform GitLab and GitHub have a large set of functionalities to further support code documentation and public releases.
- Files can be disclosed to the public, becoming a great integration of your CV, showing what you are able to do in an open and accessible way.

- Git helps to co-develop a code, test its functions and the compatibility of the various code sections.
- A long list of further possibilities became possible by git.
- Different software integration on development platforms, based on git, will help you to develop and co-develop your code.
- The platform GitLab and GitHub have a large set of functionalities to further support code documentation and public releases.
- Files can be disclosed to the public, becoming a great integration of your CV, showing what you are able to do in an open and accessible way.

- Git helps to co-develop a code, test its functions and the compatibility of the various code sections.
- A long list of further possibilities became possible by git.
- Different software integration on development platforms, based on git, will help you to develop and co-develop your code.
- The platform GitLab and GitHub have a large set of functionalities to further support code documentation and public releases.
- Files can be disclosed to the public, becoming a great integration of your CV, showing what you are able to do in an open and accessible way.

How does it work -in short-

Mary's Repository



Why should I care?

As the open libraries are exploding in numbers, you might need some criteria to assert the reliability of a project.

Unit test driven development!

That is taking full advantage of python object oriented structure.

Community

Good project are not only used by communities, but also supported

Git allows the development of projects without a clear lead. Community engagement is generally a desirable target to help develop to directly integrate feedbacks by users (and fix bugs).

Why should I care?

As the open libraries are exploding in numbers, you might need some criteria to assert the reliability of a project.

Unit test driven development!

That is taking full advantage of python object oriented structure.

Community

Good project are not only used by communities, but also supported

Git allows the development of projects without a clear lead. Community engagement is generally a desirable target to help develop to directly integrate feedbacks by users (and fix bugs).