Parallelization of TMVA in ROOT

Omar Zapata

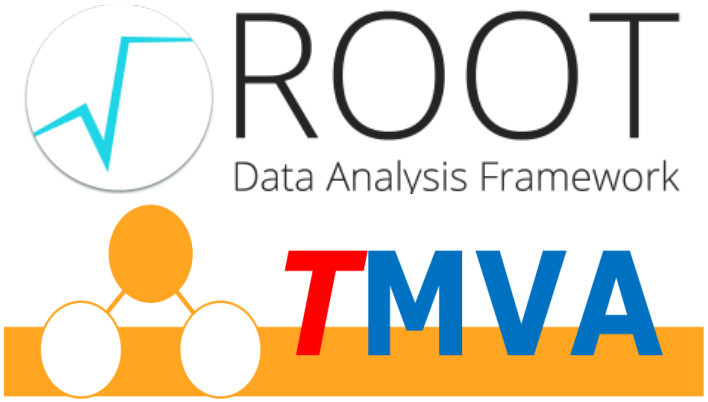
CERN Summer Student Non-Member State Program

Advisors

Lorenzo Moneta

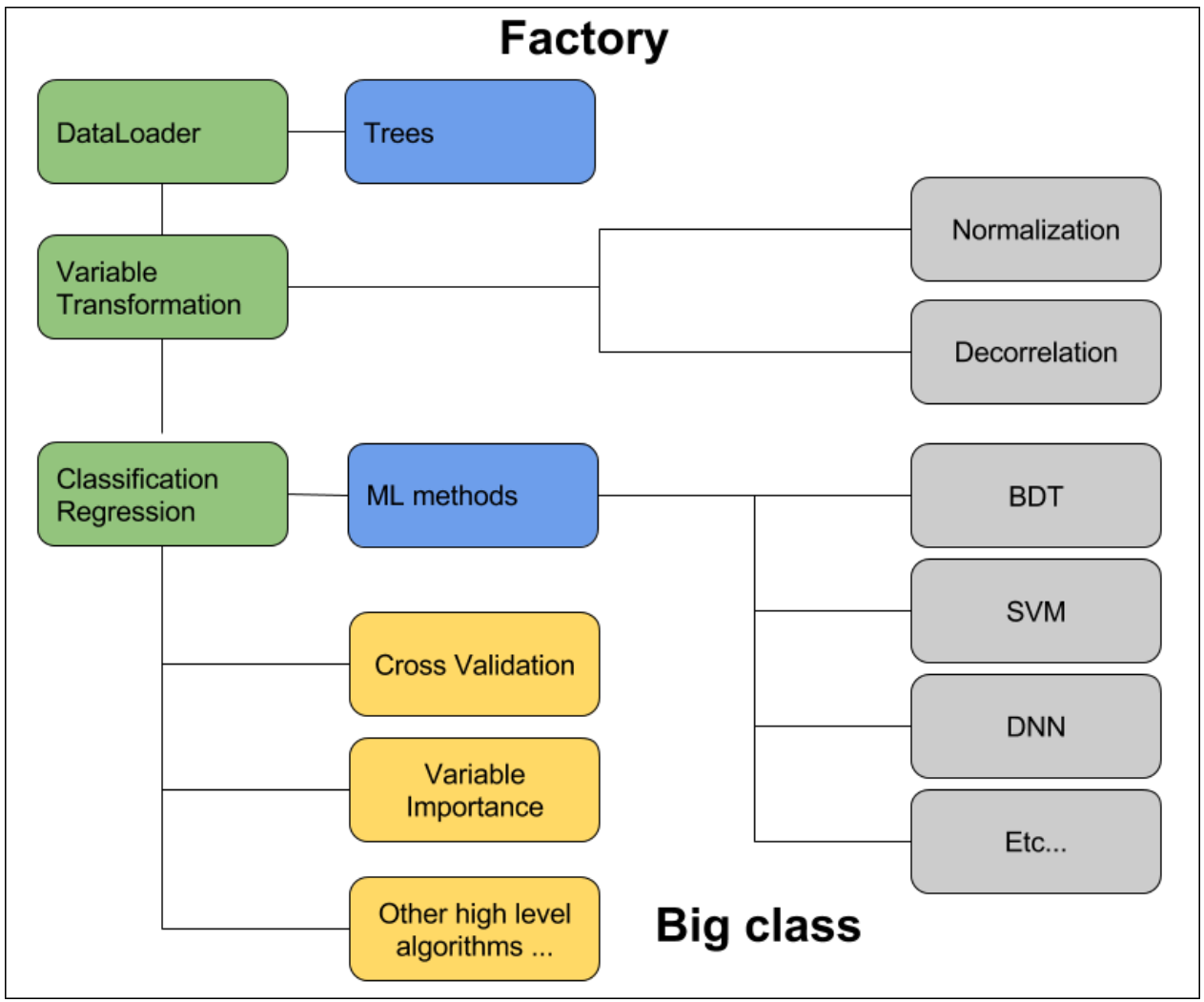
Sergei Gleyzer

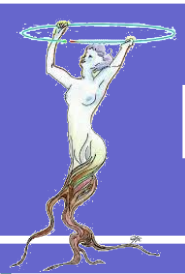
Outline



- TMVA
  - Current status
  - New architecture
  - Parallelization
- Examples
- Future outlook
- Conclusions

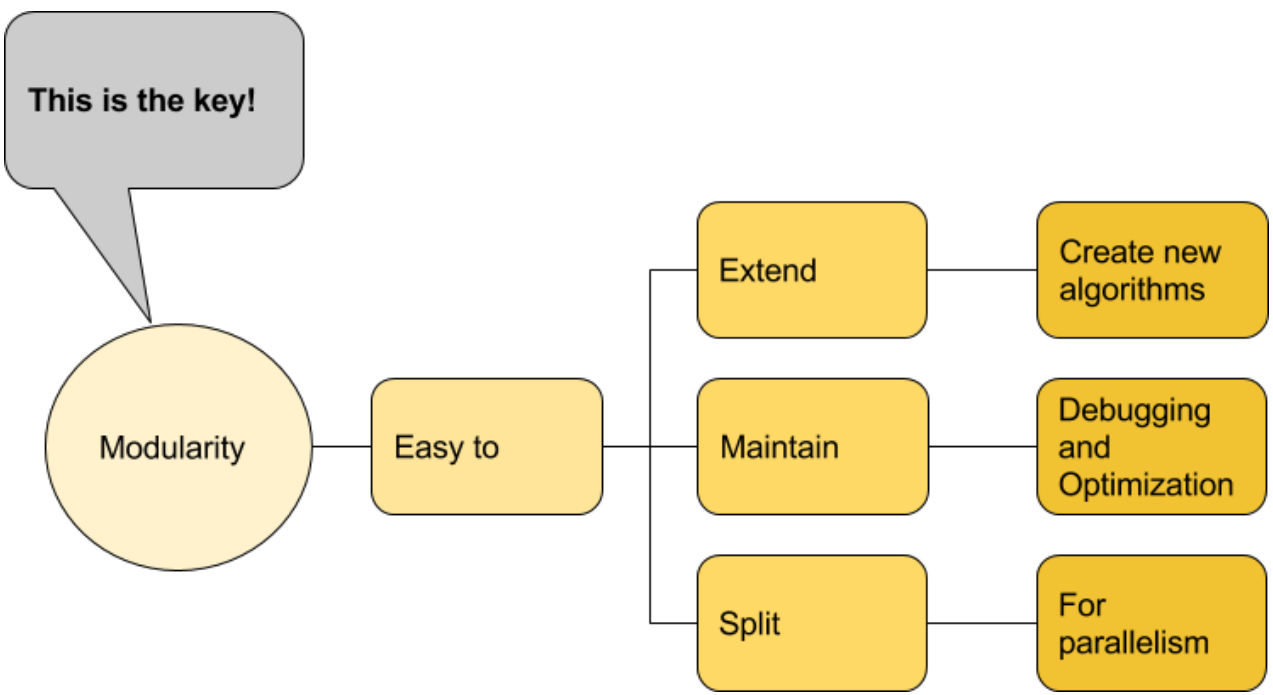
Current status





New architecture for TMVA

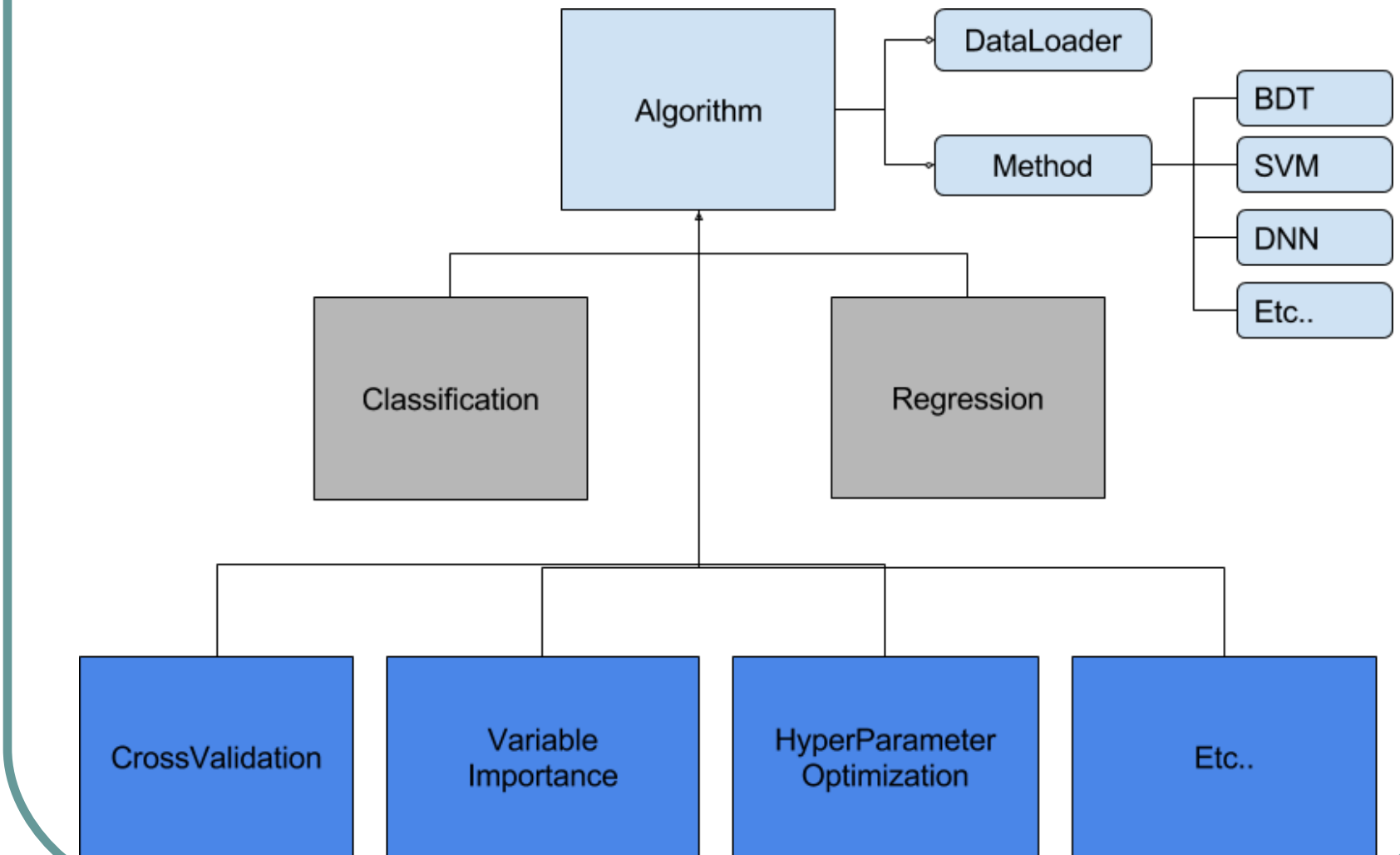
- Why?
  - Need to:
    - Parallelize the algorithms
    - Create more algorithms
    - Modularize the Factory



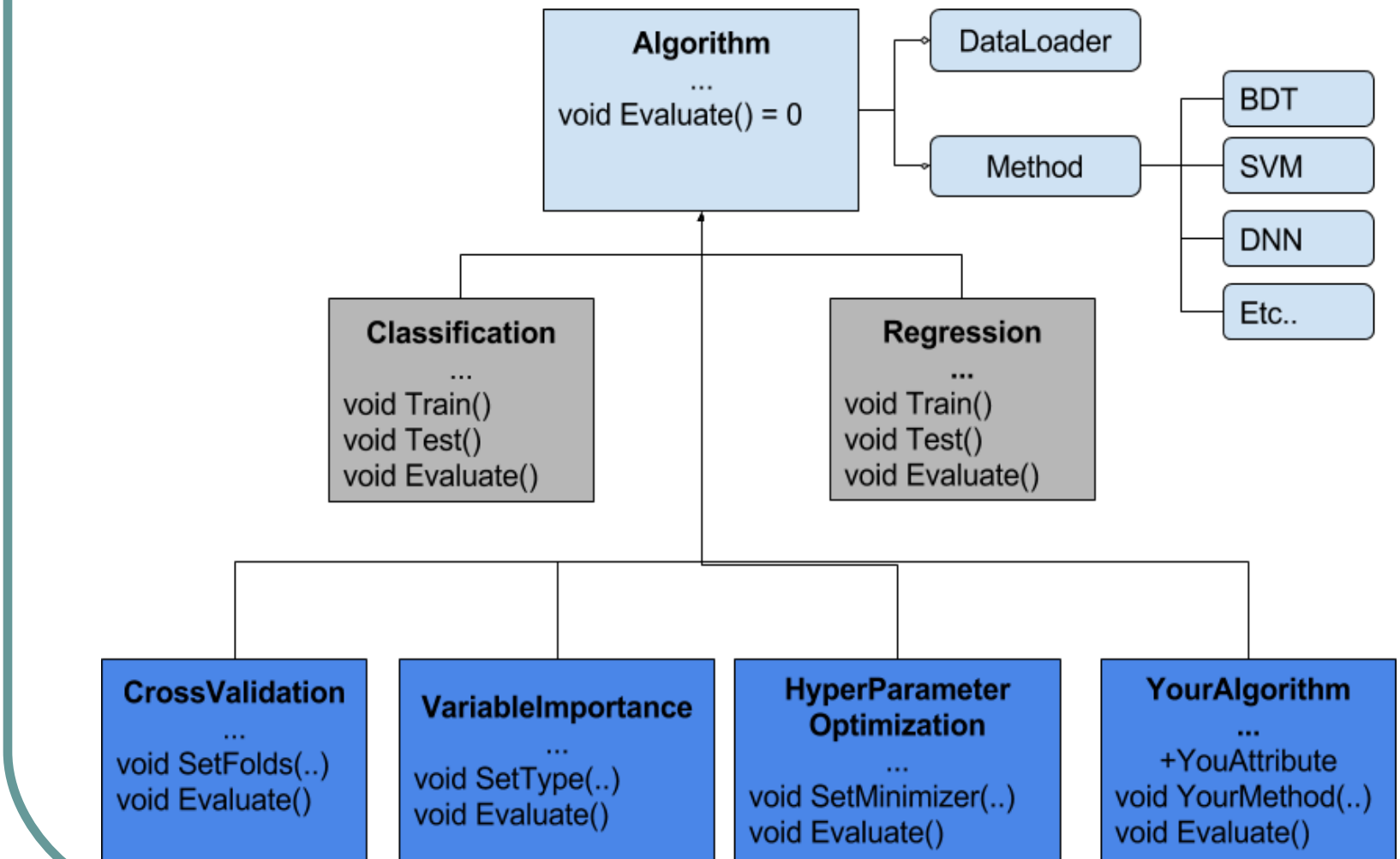
New architecture for TMVA

- Every machine learning algorithm needs:
  - Data
  - Method
  - Evaluate
- A good programming model design uses:
  - Inheritance
  - Polymorphism

New architecture for TMVA



New architecture for TMVA



Prototype with examples

Jupyter notebooks

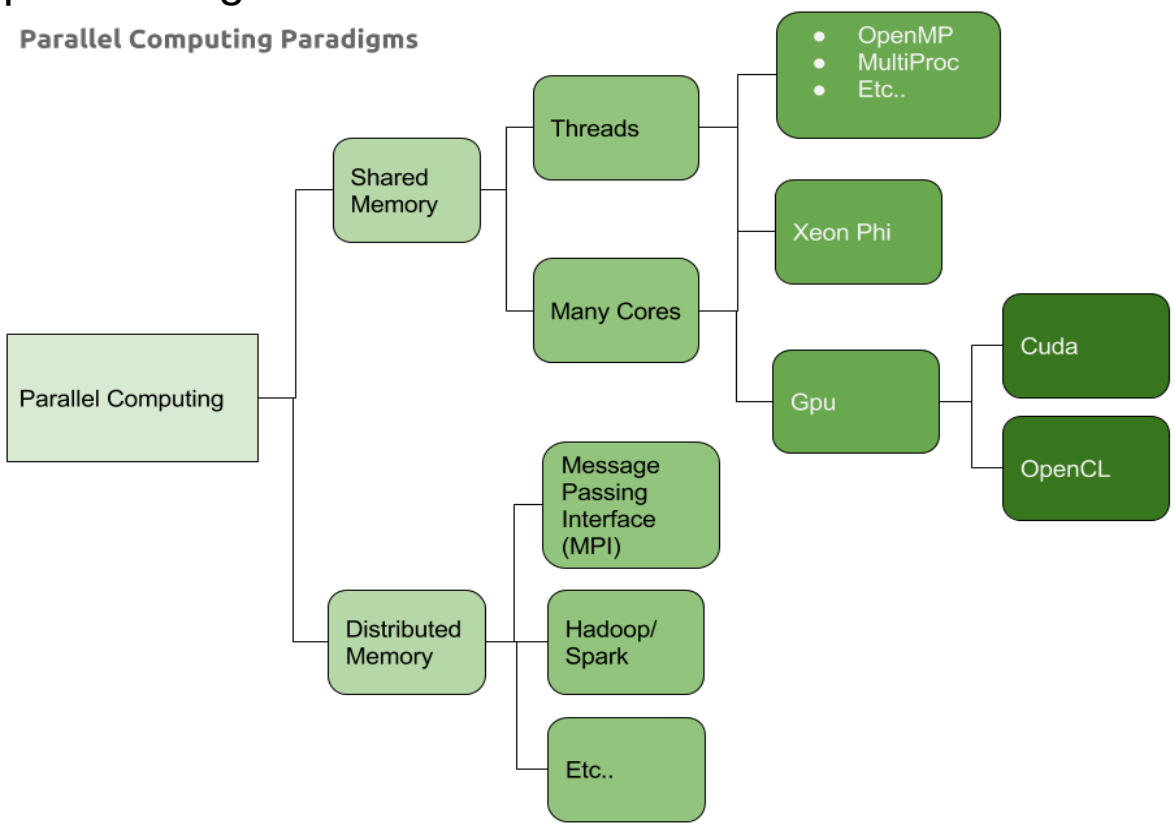
Classification

Cross Validation

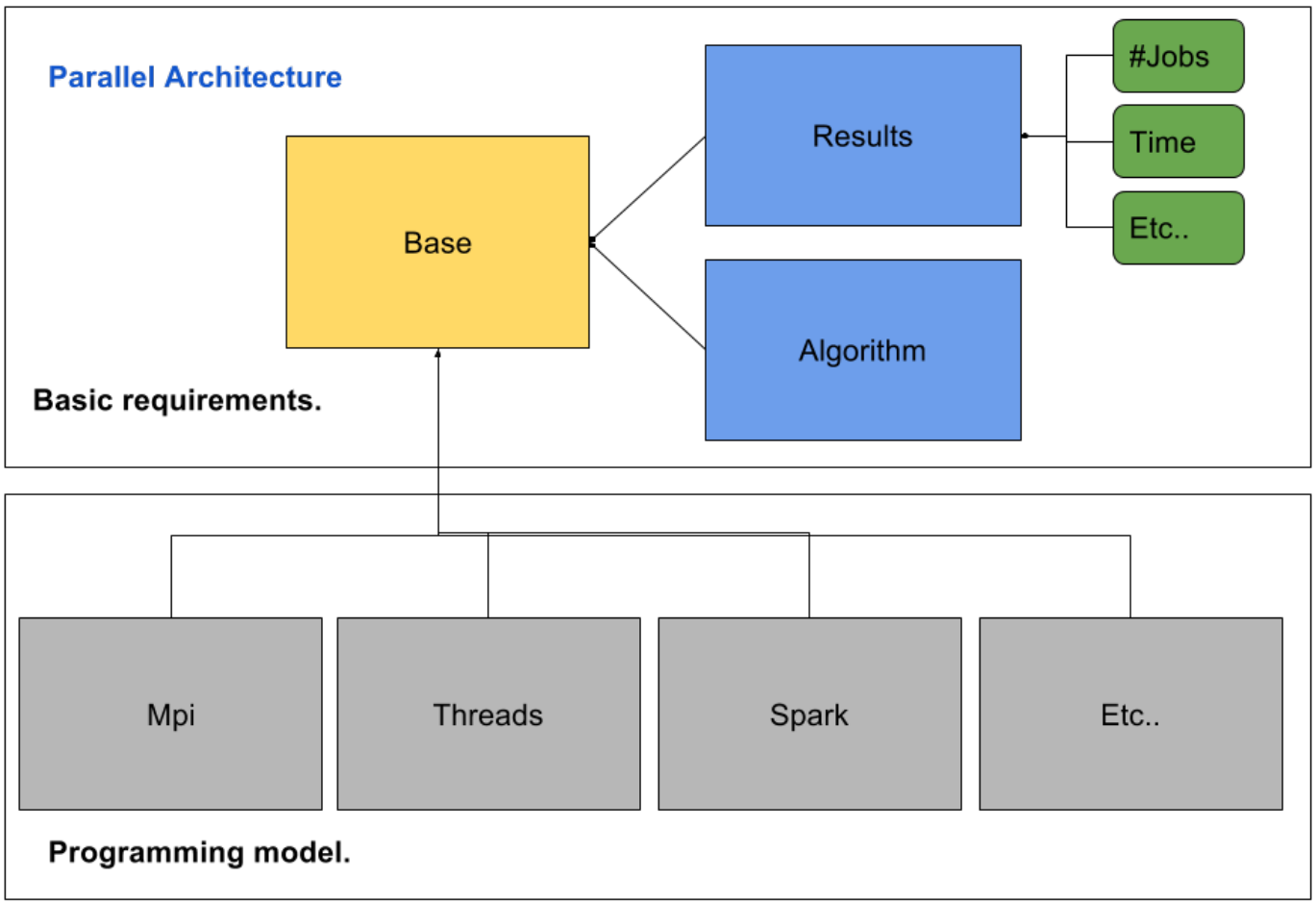
Variable Importance

Parallel architecture for TMVA

- Why parallel computing
  - Huge amount of data
  - Expensive algorithms
- Which paradigm
  - Many



New parallel architecture for TMVA



Prototype with examples

Jupyter notebooks

ParallelExecutor (MultiProc)
ParallelExecutorMpi (OpenMPI)

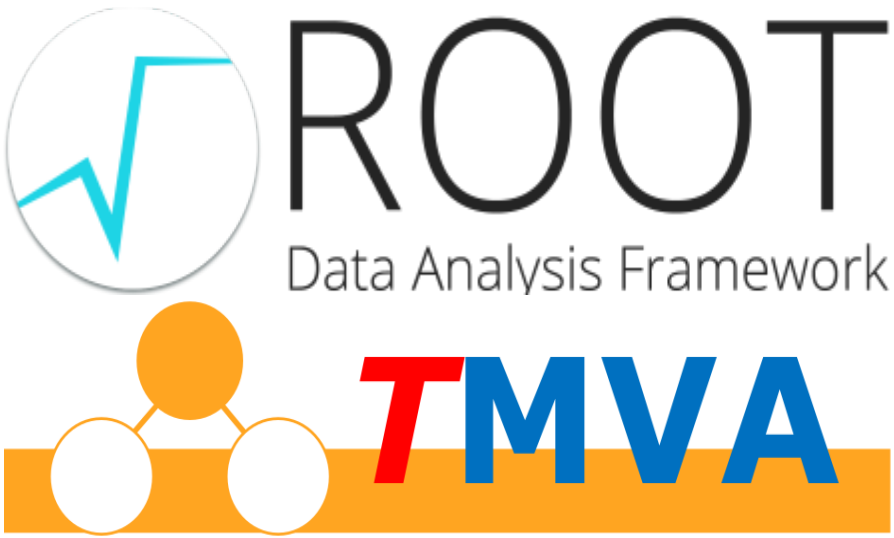
Future outlook

- In general TMVA needs
  - ROOT plugins system
  - Doxygen documentation
  - More tests for ctest
  - New features from new c++ standards
- Parallelization
  - Complete removal of static variables
  - Serialize more needed classes.
- Integrate new TMVA features.

Conclusions

- Modularity in TMVA
- Parallel architectures
  - Theads
  - MultiProc
  - Spark
  - Mpi
  - Gpu
- Great summer at CERN

More Information



Website

http://oproject.org

To finish

Thanks !