AstroData
Davide Giannuzzi, Auticon

Finanziato
dall'Unione europea
NextGenerationEU

Ministero
dell'Università
e della Ricerca

Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

Project Overview

Consortium:
Auticon SrI (Milan),
Alkemy SpA (Milan)
Net Service (Cagliari, Cosenza, Lecce)

Consultants:
University of Cagliari, Department of Physics

Finanziato
dall'Unione europea
NextGenerationEU

Ministero
dell'Università
e della Ricerca

Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

Project Overview

Main Topic
Scientific Visualization with Artificial Intelligence support

Specific Topic
Development of specialized and HPC-enabled algorithms for applying machine learning solutions to specific problems in the processing, analysis, and visualization of astrophysical data

Finanziato
dall'Unione europea
NextGenerationEU

Ministero
dell'Università
e della Ricerca

Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

Technical Objectives, Methodologies and Solutions

Research and development of a comprehensive approach to the use of Machine Learning (ML) and Deep Learning (DL) in the astrophysical field, in order to reduce technological and methodological barriers to experimentation and facilitate the use of these techniques with complex and heterogeneous data

The concrete development of a software library that implements its main components and will adopt elements from the world of HPC and Data Engineering for Big Data

Application of the framework to one specific, real-world problem

Technical Objectives, Methodologies and Solutions

We are going to use a **Design Thinking Approach** with 3 main phases:

- **Understand:** Look into the specific domain to detect problems, requirements with a multidisciplinary team and a strong governance to help achieving goals and make decisions

- **Build:** define, plan and implement efficiently MVPs (Minimum Viable Products) with an interactive approach to collect feedbacks and improve the products

- **Evolve:** support the adoption, understanding how to evolve the solution and measure the efficiency

Partners will work together with INAF to define the requirements, plan and interactly build this products, also by understanding, using and evolving INAF already done products and solutions

Technical Objectives, Methodologies and Solutions

The methodology that will emerge will be an abstraction of the main steps necessary in a generic ML project (e.g., data ingestion, data processing, ML training, data visualization), and the framework will be built using an API-based architecture, with an API for each block of the pipeline. It will be modular and extendable in any case

In the development of the tool, we will use the following best practices

- DevOps: We will use DevOps principles for the implementation of the Framework into robust and reliable software. Specifically, we will utilize Version Control, Issues Tracking, and Unit and Integration testing procedures to ensure high-quality and easily maintainable code

- DataOps: We will use DataOps principles to ensure the creation of replicable, scalable, and reliable data transformation pipelines. In particular, we will employ techniques such as Data Lineage and Data Testing

- MLOps: For training Machine Learning and Deep Learning models, we will leverage MLOps techniques to ensure the execution of reliable experiments and the achievement of correct results. Specifically, we will utilize concepts such as Experiment Tracking, Feature Store, and Model Versioning

Involved Staff and new recruitments

Role	Partner
Data Science Senior Manager	Alkemy
Senior Developers	Net Service /Alkemy / Auticon
Solution Architect	Alkemy
Domain Consultants	University of Cagliari
PM	Alkemy
Governance	Auticon / Net Service / Alkemy
Designers	Alkemy
Testers	Auticon / Net Service / Alkemy

Timescale, Milestones, SAL

WP	1	2	3	4	5	6	7	8	9	10	11	12
1 – Brainstorming and Design Thinking sessions with researchers, to identify the current SOTA (methodologies, data, infrastructure, libraries used etc.)			M1.1									
2 - Detailed definition of the Framework in all its components and a contextual survey of libraries, tools, and technologies to support it					M2.1							
3 – Software development phase and release of the MVP + relative documentation							M3.1			M3.2		
4 – Application of the framework to a real world use case											M4.1	
5 - Release of the software and documentation in Open format and dissemination of the results												M5.1

Timescale, Milestones, SAL

DELIVERABLE	WP	Milestone	Month	SAL
D1 High level definition of POC and MVP's features	WP1	M1.1	3	1 (3 months)
D2 Detailed definition of Framework's methodology, components, technologies and usage guidelines	WP2	M2.1	5	2 (6 months)
D3 POC release	WP3	M3.1	7	3 (10 months)
D4 MVP release + technical documentation	WP4	M4.1	10	
D5 Integration in the MVP of methods necessary to solve the real-world use case	WP5	M5.1	11	4 (12 months)
D6 Final Open Source release and Report on dissemination activities	WP5	M5.2	12	