Inria Brasil

A PARTNERSHIP WITH LNCC AND BEYOND

Building subset models for short term forecast of extreme events

AI for extreme events

INRIA- BRASIL: Workshop on Digital Agriculture

Fabio Porto

fporto@lncc.br

Instituto de Inteligência Artificial

Laboratório Nacional de Computação Científica

UNIDADE DE PESQUISA DO MCTI

DEXL

DATA EXTREME LAB

Agenda

- Extreme Events Challenges
- 4 steps Approach
- Subset Models
- Final Comments

Fabio Porto - fporto@lncc.br

Extreme Events: Rainfall in Brazil

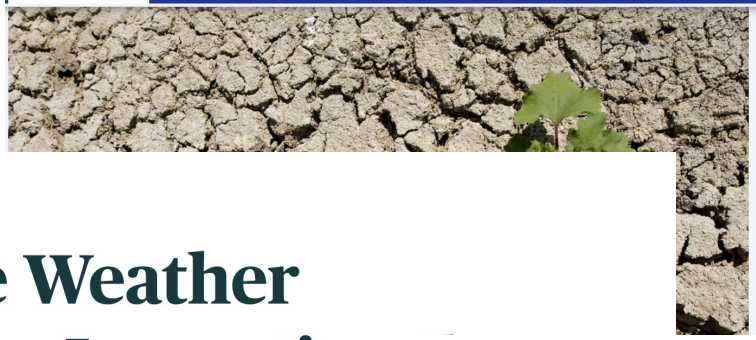
Fabio Porto - fporto@lncc.br

EE and Agriculture in the rest of the world

eco logic

EN | DE

Topics Projects Publications Events Presentations Team Ins



The Impact of Extreme Climate Events on Agricultural Production in the European Union

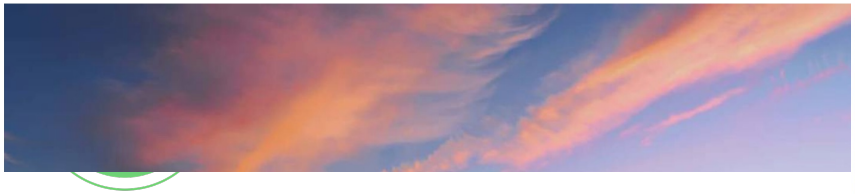
PROJECT

DURATION Nov 2022 - Feb 2023

2 MARS 2023

How Extreme Weather Conditions Are Impacting Agrifood Chain?

Author: Ali NASRALLAH / Article / Topics : Agriculture



AGRICULTURAL RECRUITMENT SPECIALISTS

Find Jobs

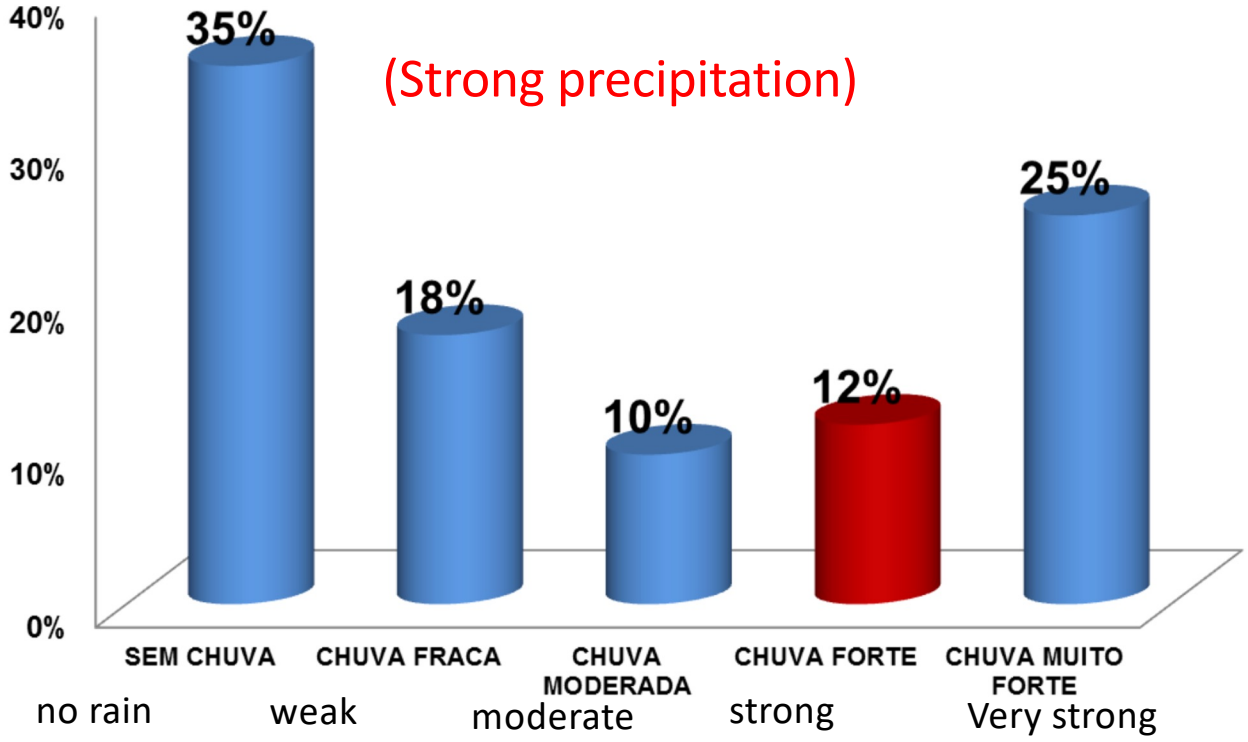
Home Job Search Candidates Client Services Executive Search About Us Work For Us Insights

Forecast x Observation

02/03/2020 – Rio de Janeiro - 4 deaths

Padre Miguel	220,2 mm
Santa Cruz	205,2 mm
Realengo	191,0 mm
Bangu	182,6 mm
Alto da Boa Vista	181,8 mm
Anchieta	171,2 mm
Vargem Pequena	163,8 mm
Jacarepaguá	158,6 mm
Grota Funda	157,4 mm
Barra/Barrinha	152,8 mm

Forecast x Observation



(Strong precipitation)

Alerta Rio System

Source: Fabricio Polificke da Silva, Final report, Alerta Rio, 2019

https://rionowcast.dexl.lncc.br/

RIO NOWCAST

Home About Publications Subprojects Slides News Events Datasets

Highlights News : Projeto de inteligência artificial pode prever desastres climáticos

Rionowcast

"A research and development project to improve extreme rainfall forecast using meteorological data and artificial intelligence models."

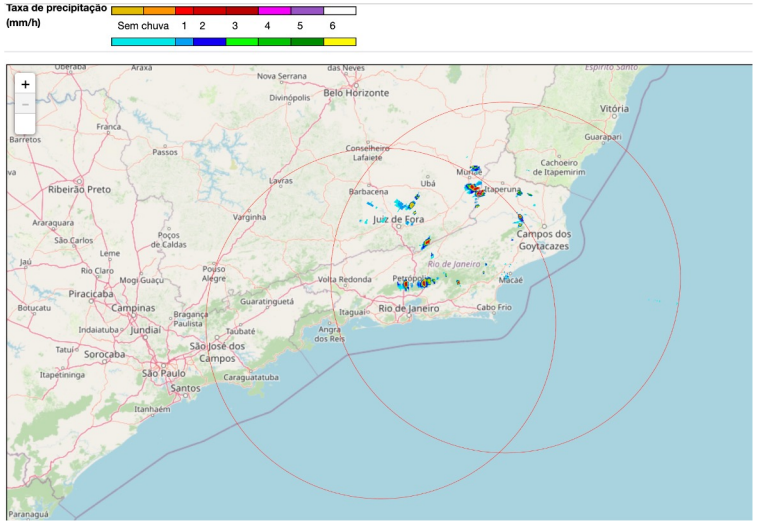
Instituto de Inteligência Artificial

DEXL DATA EXTREME LAB

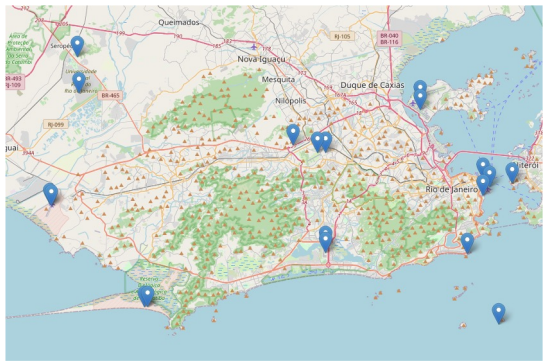
Relevant data sources

Numerical simulations (GFS-25, WRF), Reanalysis (ERA-5 ECMWF)

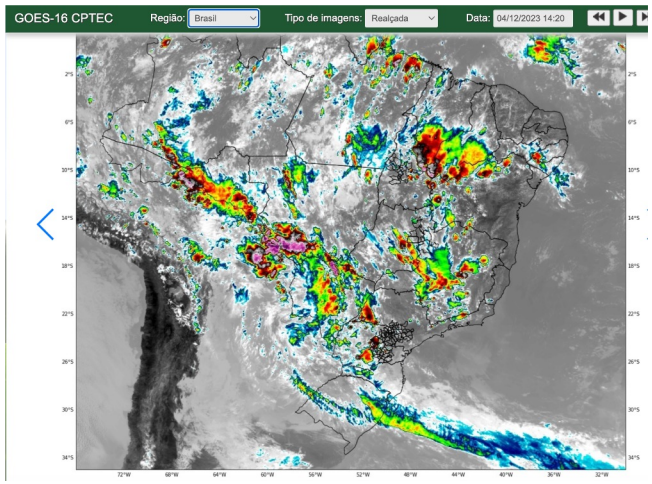
Ocean Bouys



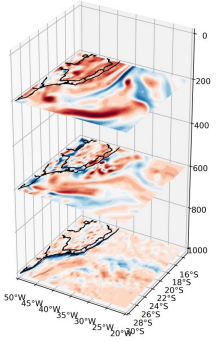
Radars: Guaratiba, Macaé, Sumaré



Rain gauges / Weather station



Satellite: Goes 16

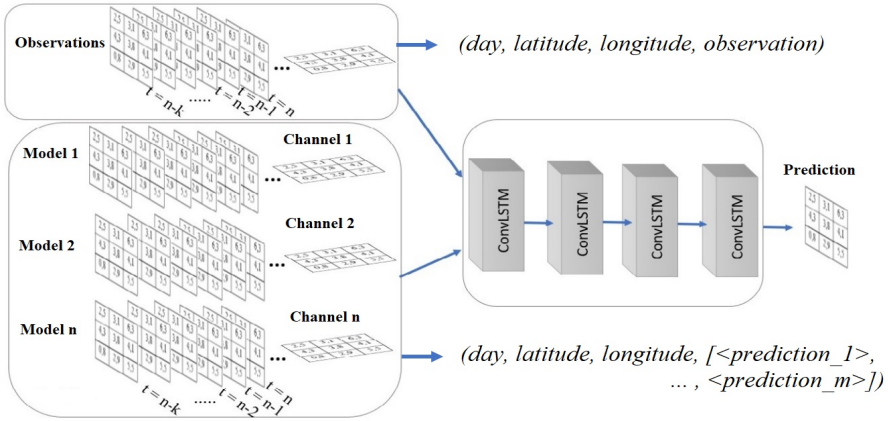


ERA5

SBGL (radiosonde)

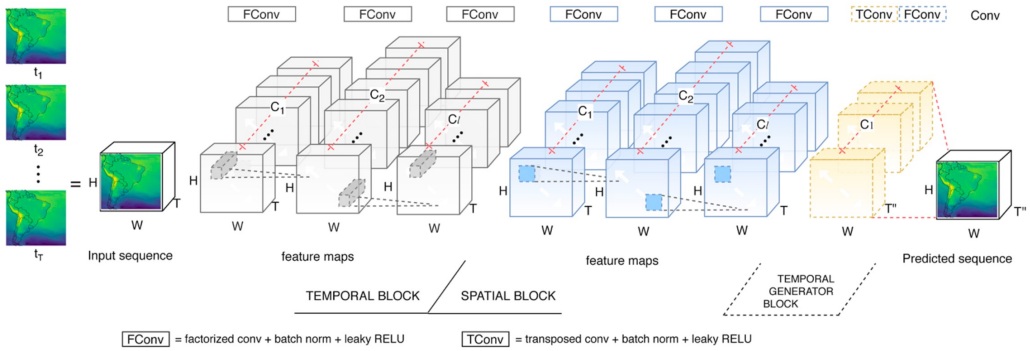
Rainfall Prediction Models

YCONVLSTM



Source: Yania Souto et al., A Spatio-temporal ensemble approach to Rainfall Forecasting, IJCNN, 2018

STConvS2S architecture



source: Nascimento,R.C. et al, STConvS2S: Spatiotemporal Convolutional Sequence to Sequence Network for Weather Forecasting, Neurocomputing, 2021

Challenges

- Complex and chaotic phenomenon
- Heterogeneos data sources
- Continous and voluminous data
- Extremely unbalanced data distribution (no rain 90%)
- Numerical Models : frequent data assimilation
- Spatio-temporal ML models: irregular and regular spatial data; unbalanced and missing data; plethora of learning algorithms

Inria

Hot topic

Laboratório Nacional de Computação Científica
UNIDADE DE PESQUISA DO MCTI

RESEARCH

GraphCast: AI model for faster and more accurate global weather forecasting

14 NOVEMI

Remi Lam on behalf of

NVIDIA

Home | AI | Data Center | Driving | Gaming | Pro Graphics | Robotics | Healthcare | Startups | AI Podcast | NVIDIA Life

AI Chases the Storm: New NVIDIA Research Boosts Weather Prediction, Climate Simulation

Amid hurricane season, NVIDIA unveils breakthrough generative AI model for emulating high-fidelity atmospheric dynamics.

August 19, 2024 by Mike Pritchard

nature

Explore content | About the journal | Publish with us

nature > articles > article

Article | Open access | Published: 22 July 2024

Neural general circulation models for weather and climate

Dmitrii Kochkov, Janni Yuval, Ian Langmore, Peter Norgaard, Jamie Smith, Griffin Mooers, Milan Klöwer, James Lottes, Stephan Rasp, Peter Düben, Sam Hatfield, Peter Battaglia, Alvaro Sanchez-Gonzalez, Matthew Willson, Michael P. Brenner & Stephan Hoyer

Inteligência Artificial

Fabio Porto - fporto@lncc.br

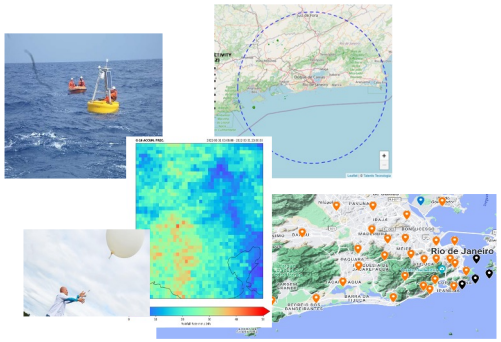
DEXL
DATA EXTREME LAB

Our Approach

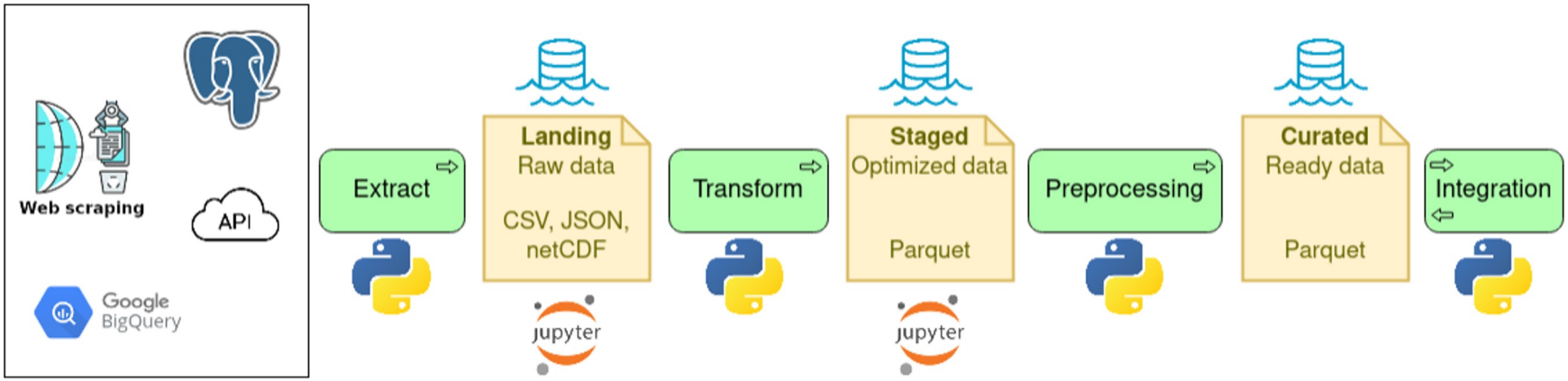
- Data Management
- Gypscie System
- Data-centric approaches for model construction
- Spatio-temporal ML models
- Data integration through Knowledge graphs

Fabio Porto - fporto@lncc.br

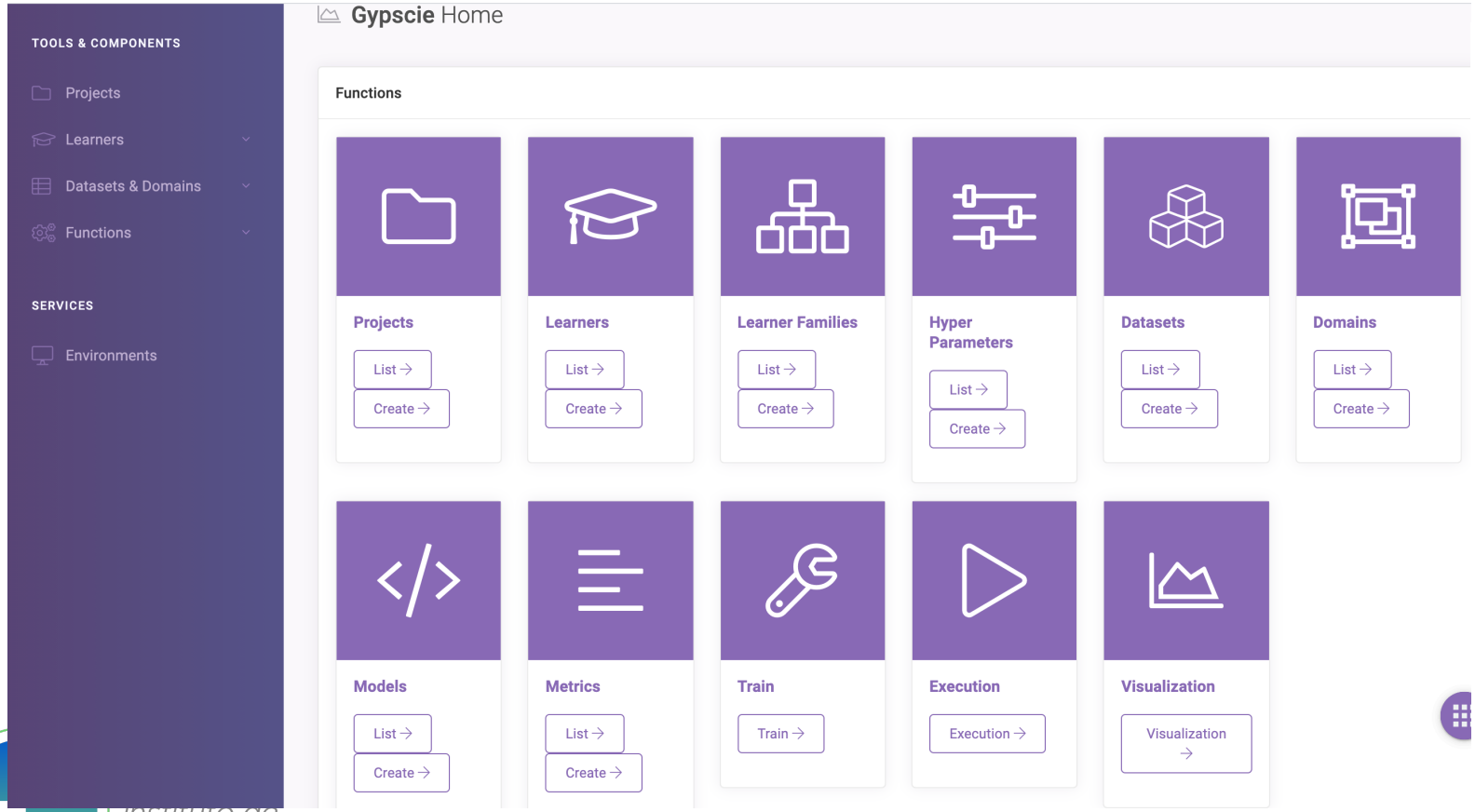
Data Management



Data Lake built with MinIO.



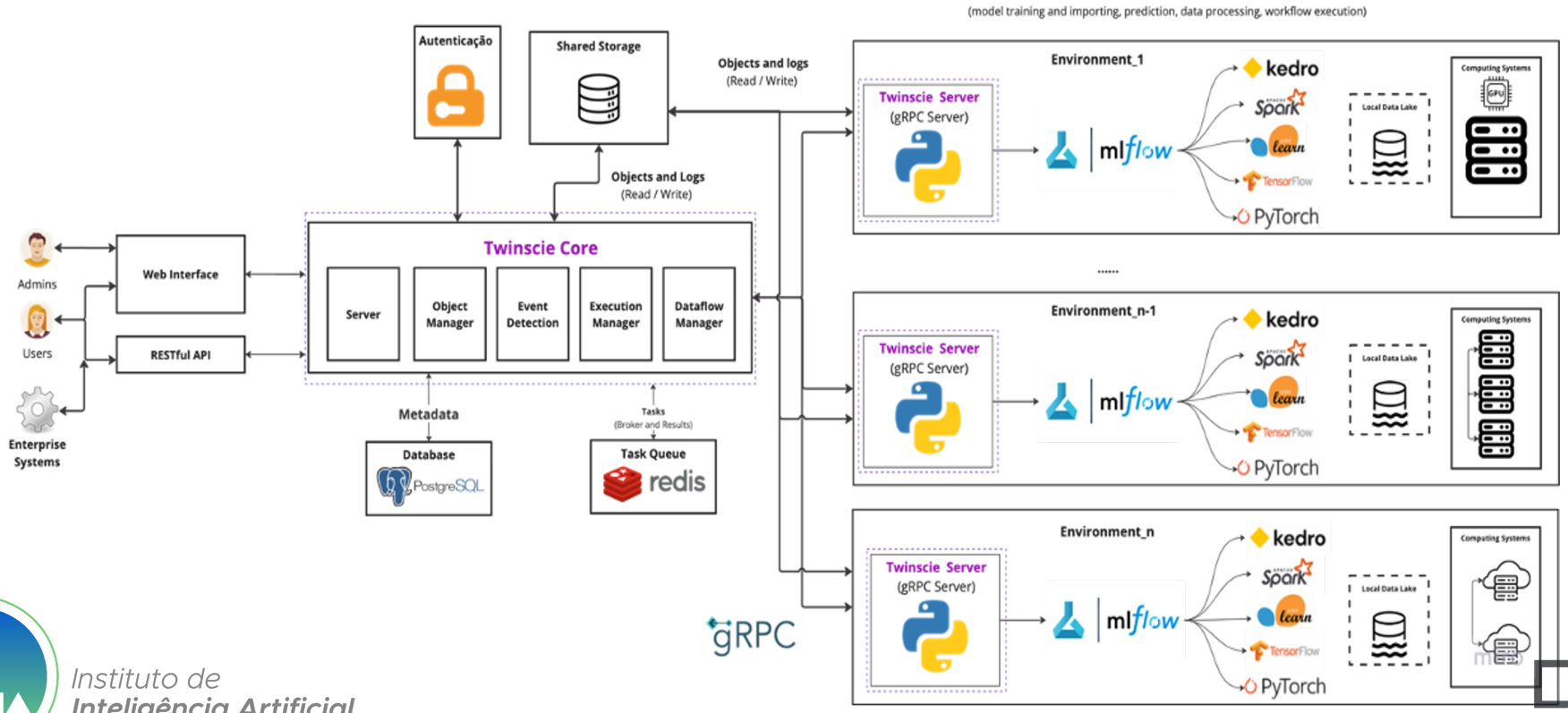
Gypscie: AI Model Management



GYPSCIE

Gypscie System

management of data and models.

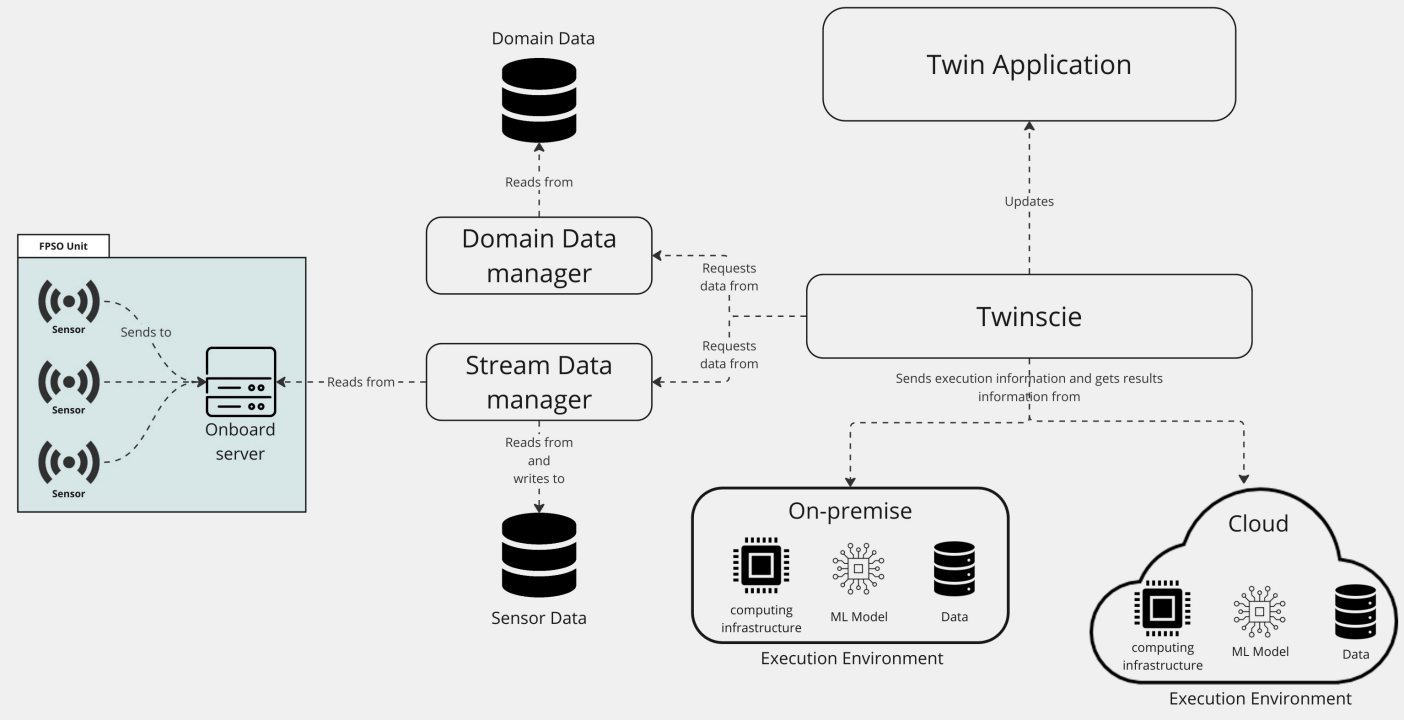


Instituto de Inteligência Artificial

Laboratório Nacional de Computação Científica
UNIDADE DE PESQUISA DO MCTI

DEXL
DATA EXTREME LAB

Digital Twin System



Instituto de Inteligência Artificial

Fabio Porto - fporto@lncc.br

DEXL DATA EXTREME LAB

Subset ML Models

- Let's asssume a dataset *D=(X,Y)* representing a phenomena ρ
  - D={p1,p2,..,pm}
    - in most of our experiments pi is a time-series at a location i
    - could be an independent sample, eg: an individual of a species
- A ML model M(X)->Y is trained in a training dataset D_{tr} having as predictors X values and as target Y values.
  - supervised learning
  - mostly regression tasks but it can be a classification problem too
- Global Model – An ML model M built using a complete dataset (D)
- Local Model – An ML model M_i built on a subset of a dataset *D*, $S \sqsubseteq D$

The case for Volume: CRAIG (coresets)

- A model is trained on a full domain coverage $\mathcal{D}$
  - for large $\mathcal{D}$ time and computational costs may become prohibitive
- Can one quickly find a subset $S \sqsubset \mathcal{D}$ s.t. the model is trained on S samples and (approx.) converges to the optimal solution (i.e. the same parameters that would have been found had $\mathcal{D}$ been used instead)
  - where should we find the samples in S: on the boundaries; the most diverse set of points ??
  - Thus, we want to find S s.t.:
  - $S^* = argmin_{S \sqsubseteq D, \gamma_j \geq 0 \; \forall j} \gamma_j |S|$ s.t.
    $max_{w \in W} \left\| \sum_{i \in D} \nabla f_i(w) - \sum_{j \in S} \gamma_j \nabla f_j(w) \right\| \leq \varepsilon.$

Mirzasoleiman, B., Bilmes, J., Leskovec, J., Coresets for Data-efficient Training of Machine Learning Models, ICML, 2020

Other ongoing approaches

- Graph Neural Nets to data interpolation using radar data.
- Evaluation of Transformed-based models for EE prediction
- Constructing a Knowledge Graph to integrate historic, current and predicted state of variables
  - data analytics; explainability; reproducibility

Final Comments

- Our Research on data + models management can be applied to Agriculture problems -> Gypscie system
- We will be working on
  - Construction of Subset ML Models
  - Efficient ML dataflows execution
  - Similarity among extreme events multi-variate time series
  - Knowledge graph construction
  - Physics constrained ML models

Fabio Porto - fporto@lncc.br

Thank You !!

fporto@lncc.br