



university of
 groningen

TNO



Bridging The Gap

Vitens



RESEARCHABLE

Empowering Machine Learning Development
with Service-Oriented Computing Principles

Mostafa Hadadian, Viktoriya Degeler, Alexander Lazovik



Evolutionary Changes in Data Analysis (ECiDA)



MOSTAFA
HADADIAN



ALEXANDER
LAZOVIK



VIKTORIYA
DEGELER



rijksuniversiteit
 groningen



UNIVERSITEIT VAN AMSTERDAM



RESEARCHABLE

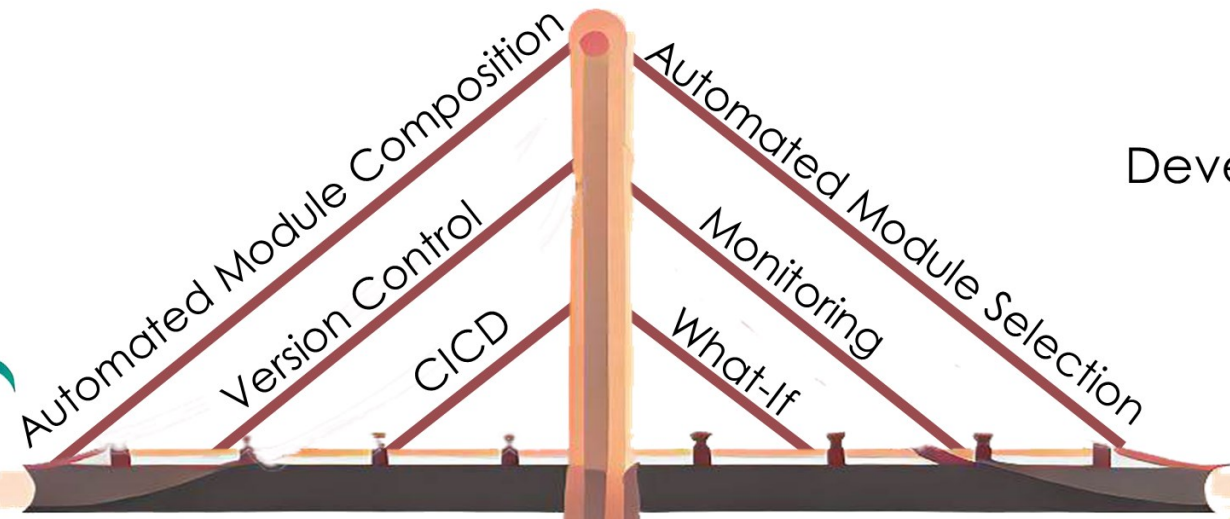


Service-Oriented Computing

Machine Learning Development

Mature
Clear Scope
Predefined Logic
Development for production

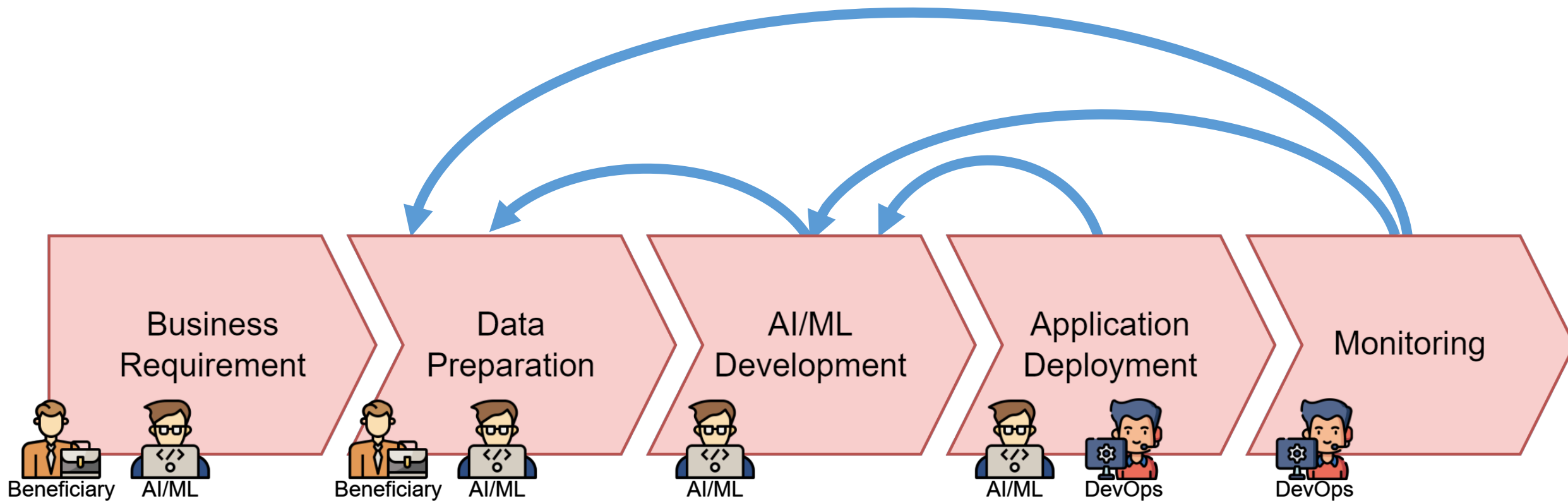
Evolving
Probabilistic
More Iterative
Evolving Scope
Data-Driven Update
Development for Exploration



**Adaptive
Runtime
Controller**

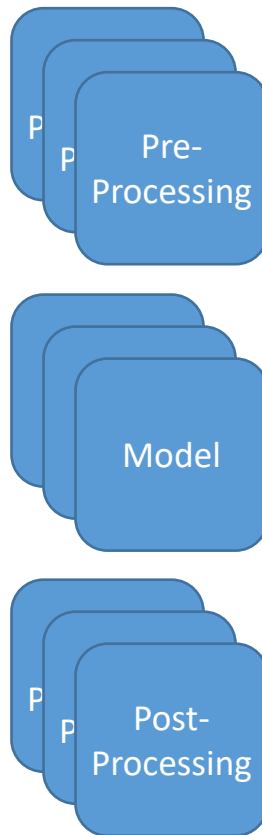
Infrastructure

Coarse-Grained AI/ML Workflow

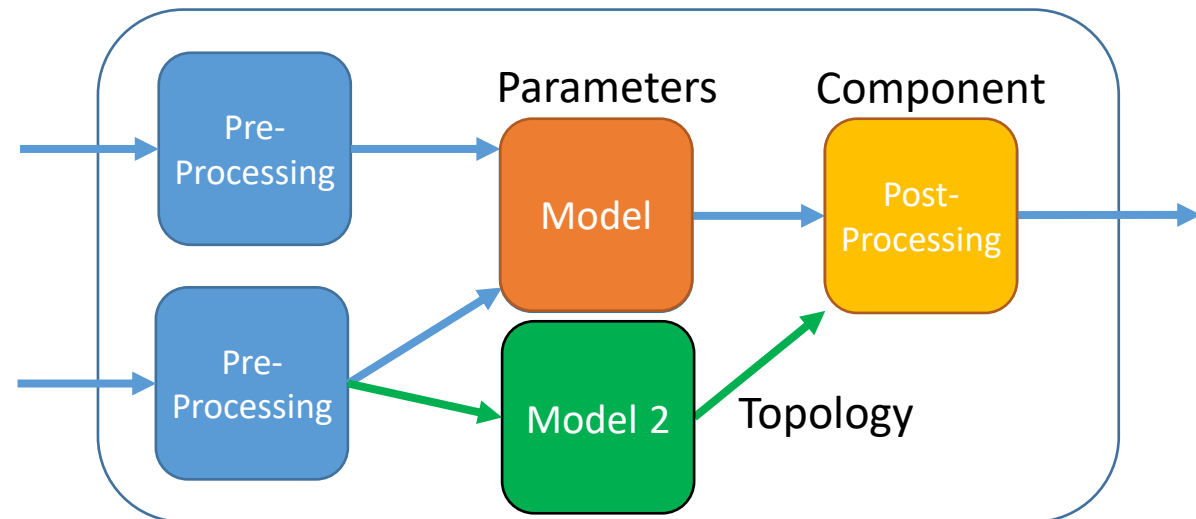


Data Processing Pipeline

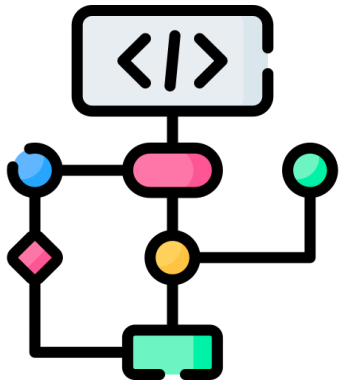
Individual Component



Pipeline

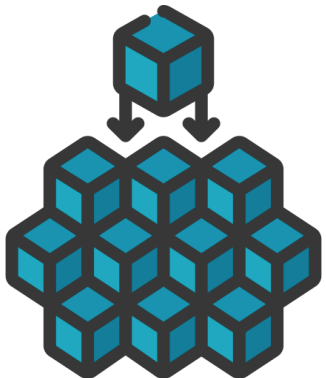


Modularity By Design



Algorithmic Modularity

- Utilization of programming languages or frameworks for the development of machine learning applications



Architectural Modularity

- Packaging each stage into distinct services and deploying these services into appropriate environments

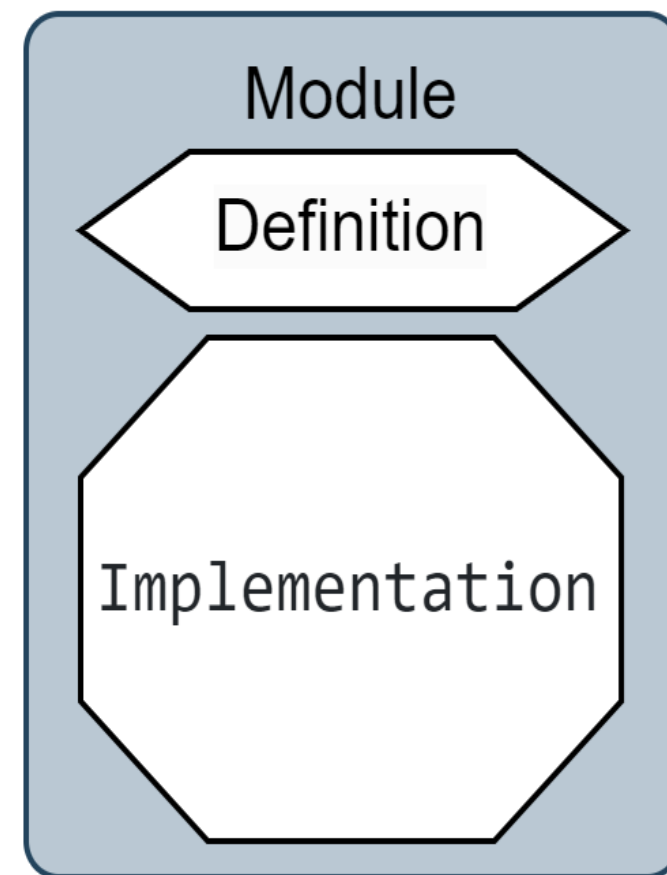
Everything as a Module (XaaM)

Module Definition

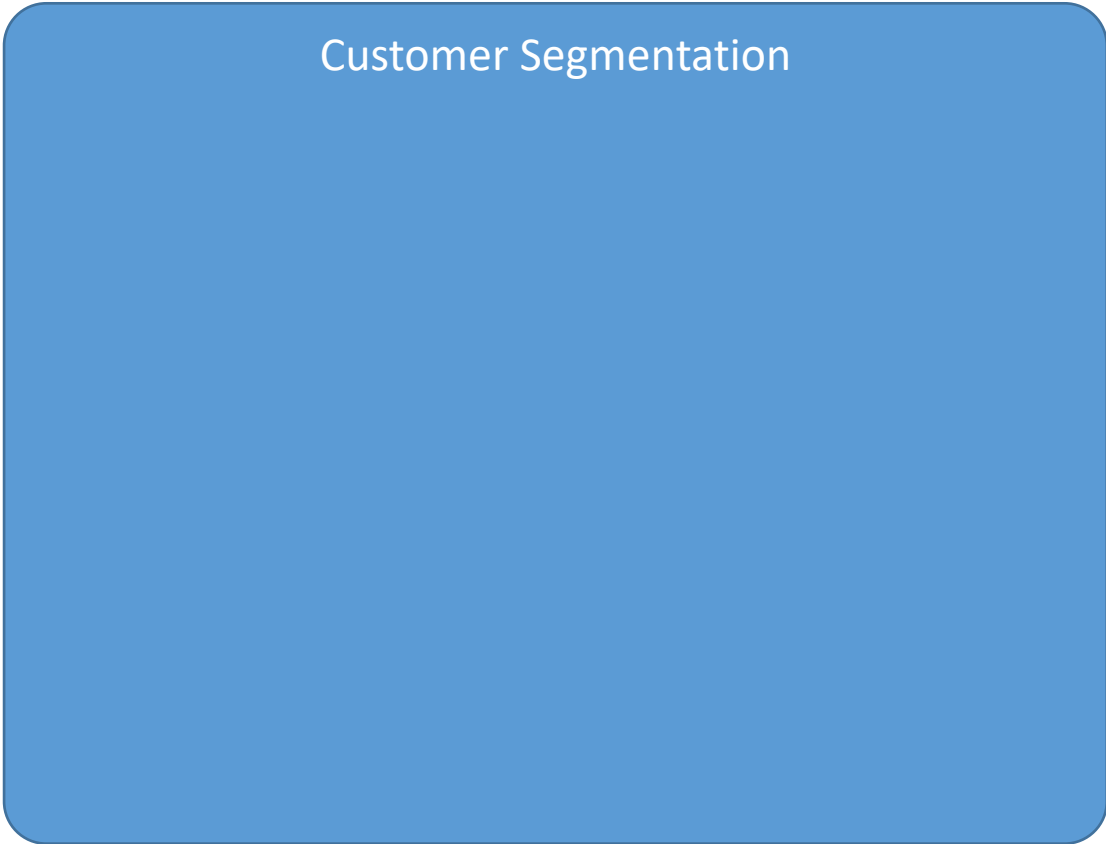
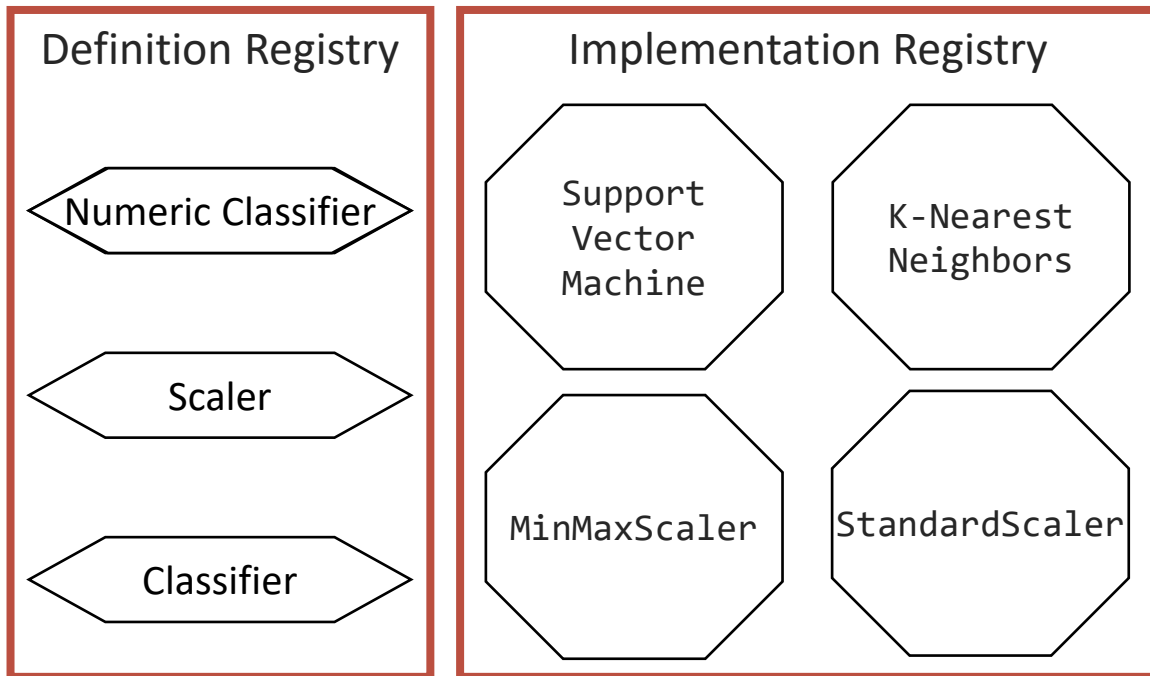
- A general and unified interface
- Separation of responsibilities
- Improved team communication

Module Implementation

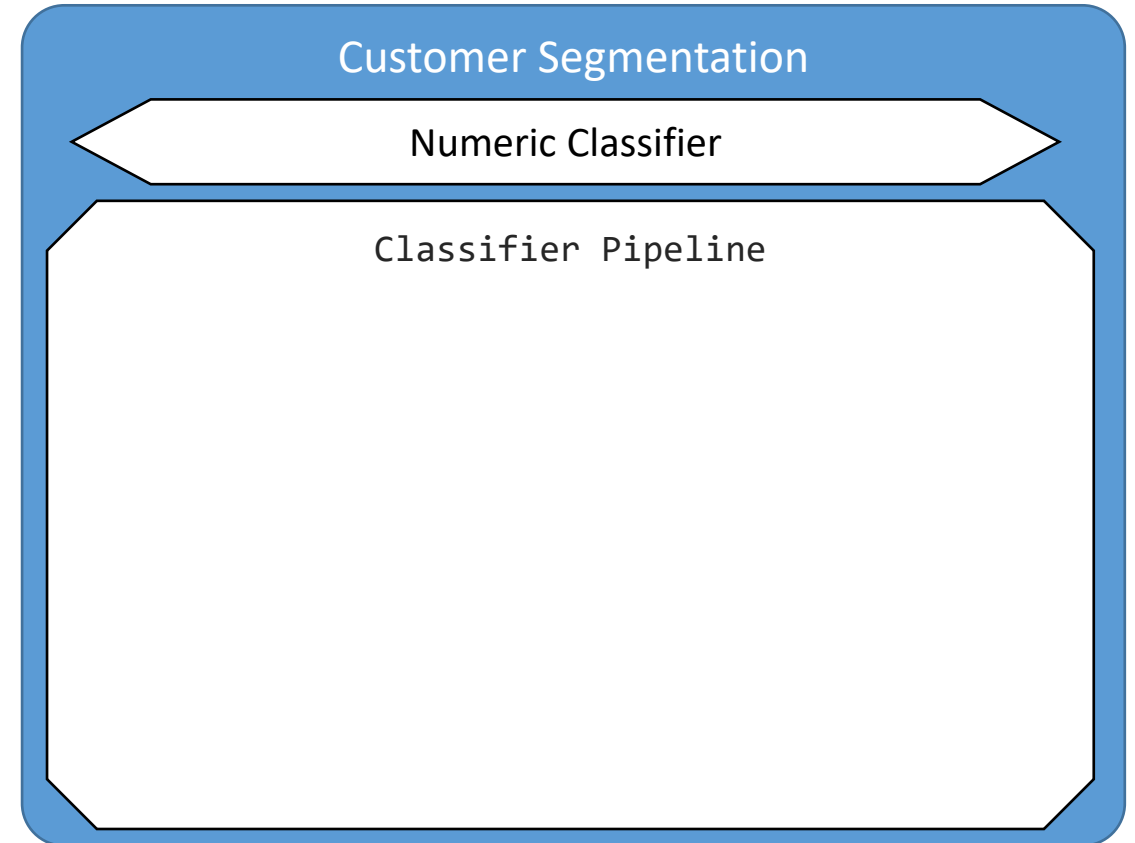
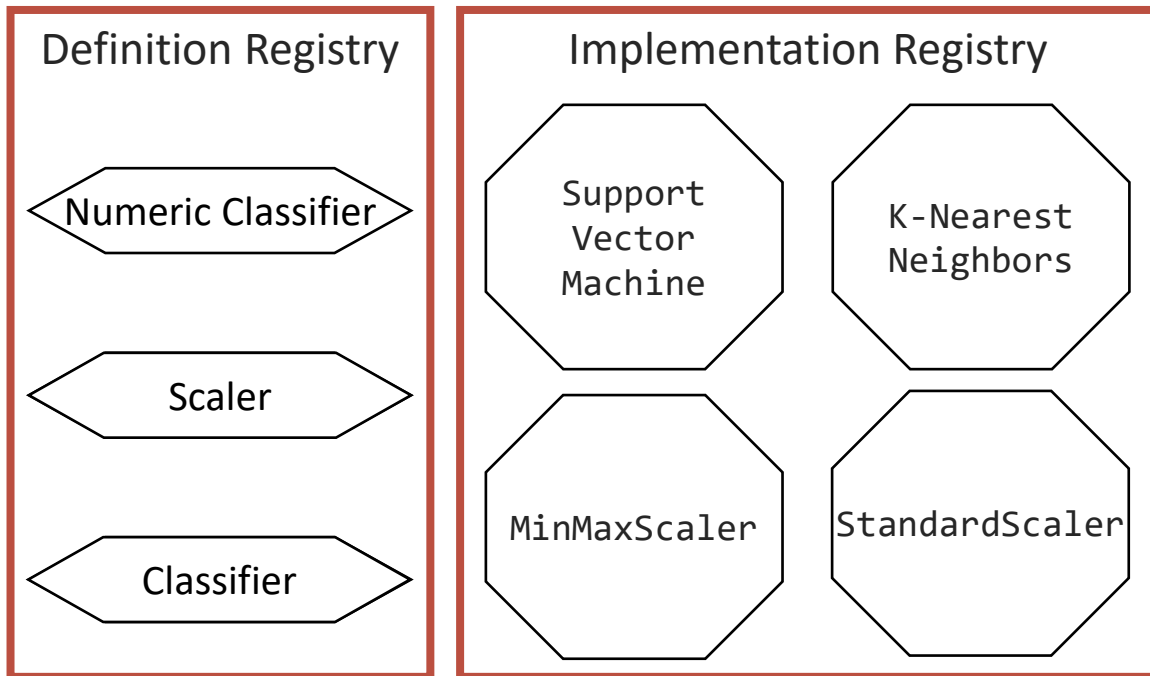
- Puts definition into practice
- Reusability and polymorphism



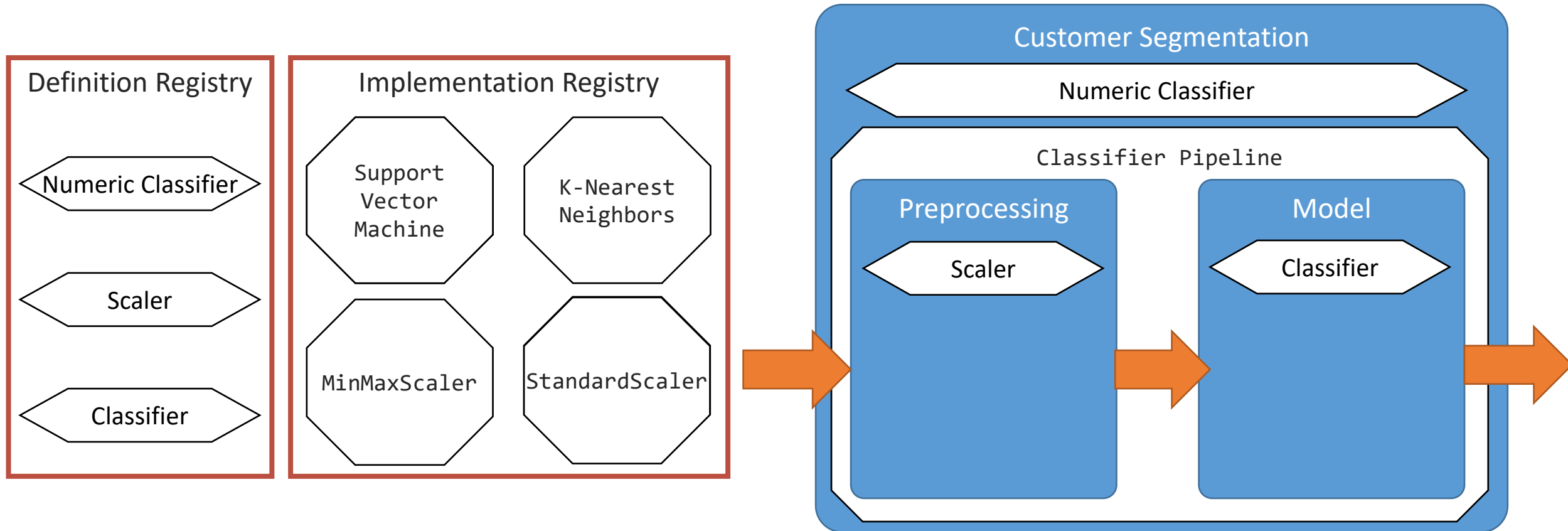
Module Composition Example



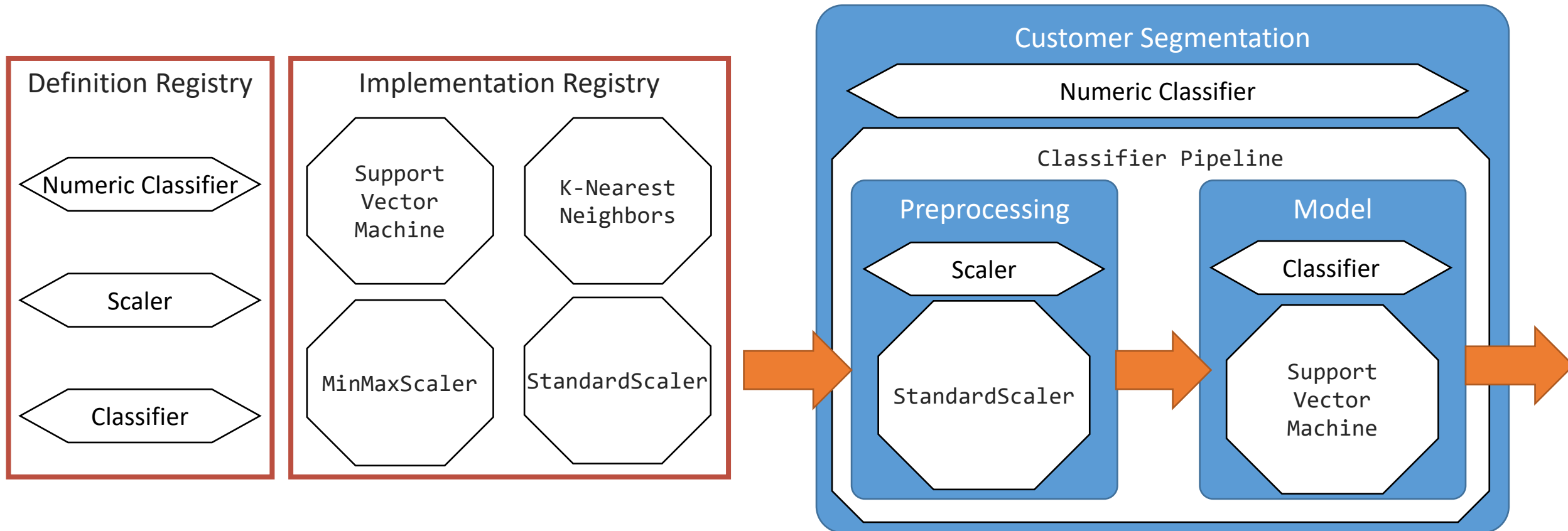
Module Composition Example



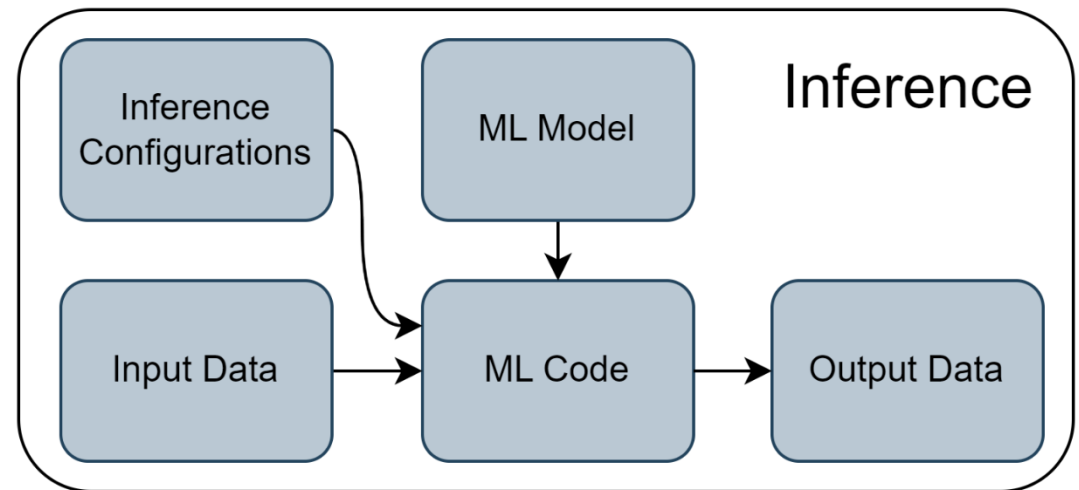
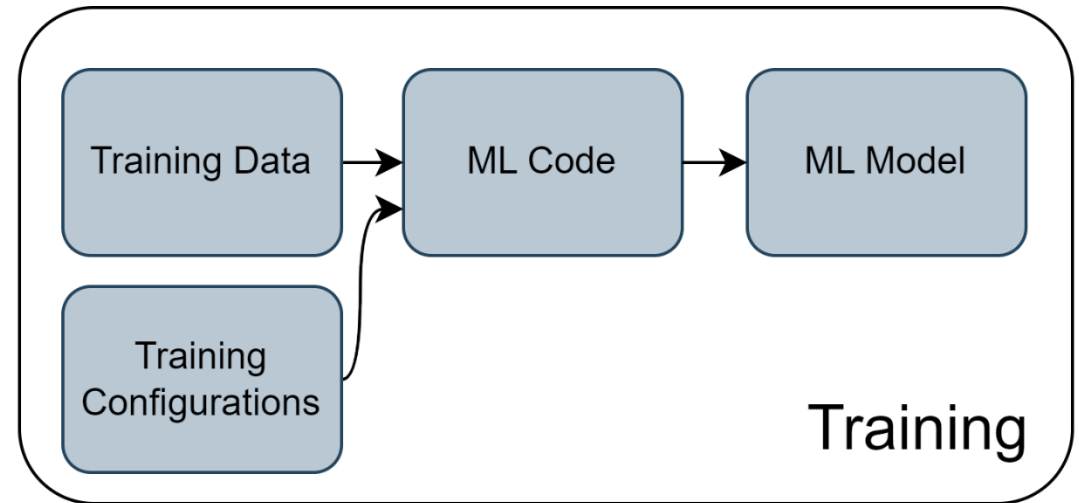
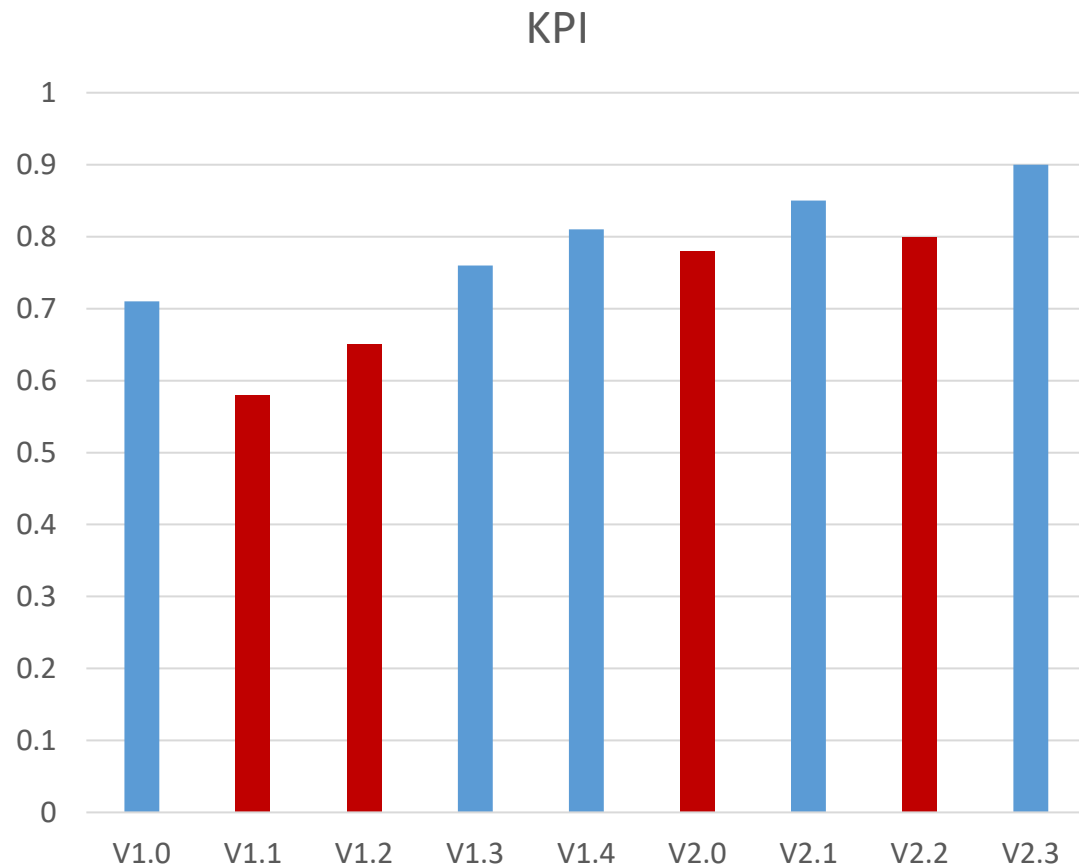
Module Composition Example



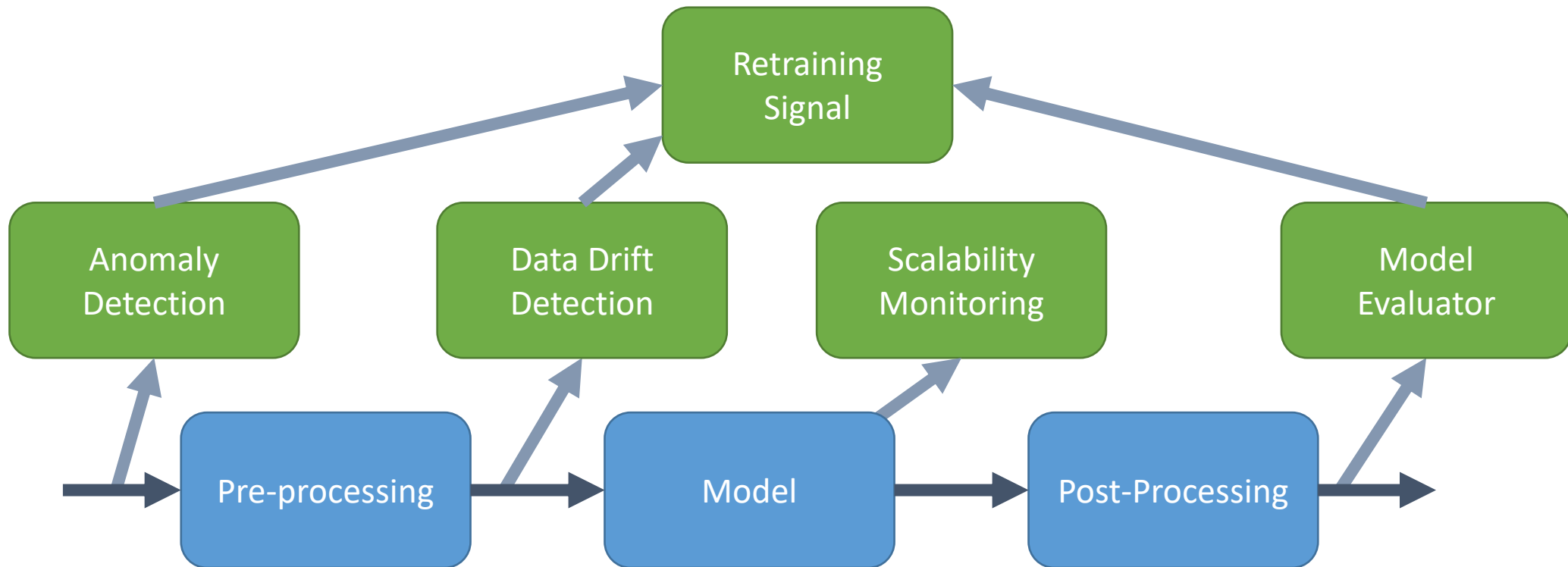
Module Composition Example



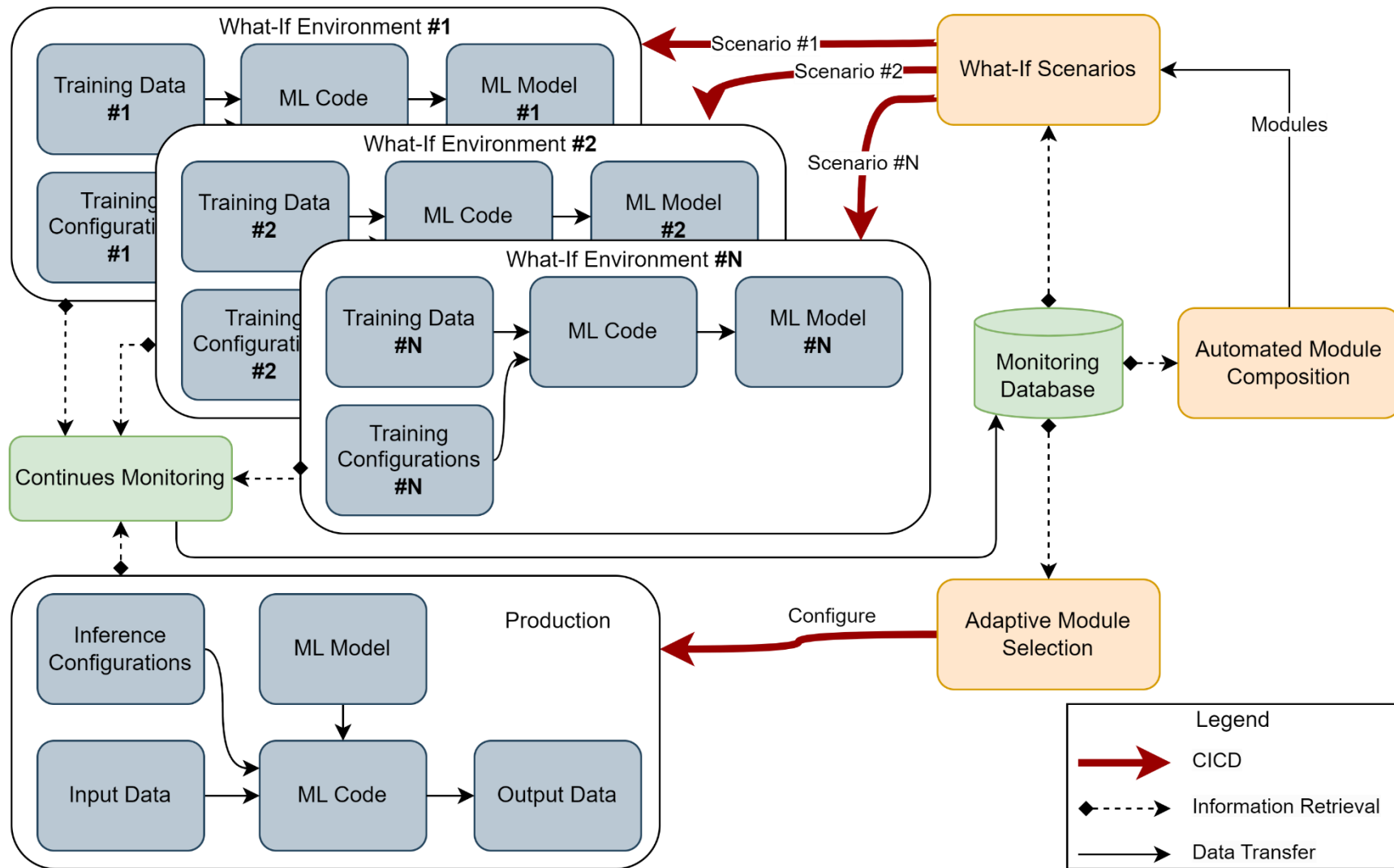
Version Control



Monitoring as a Module

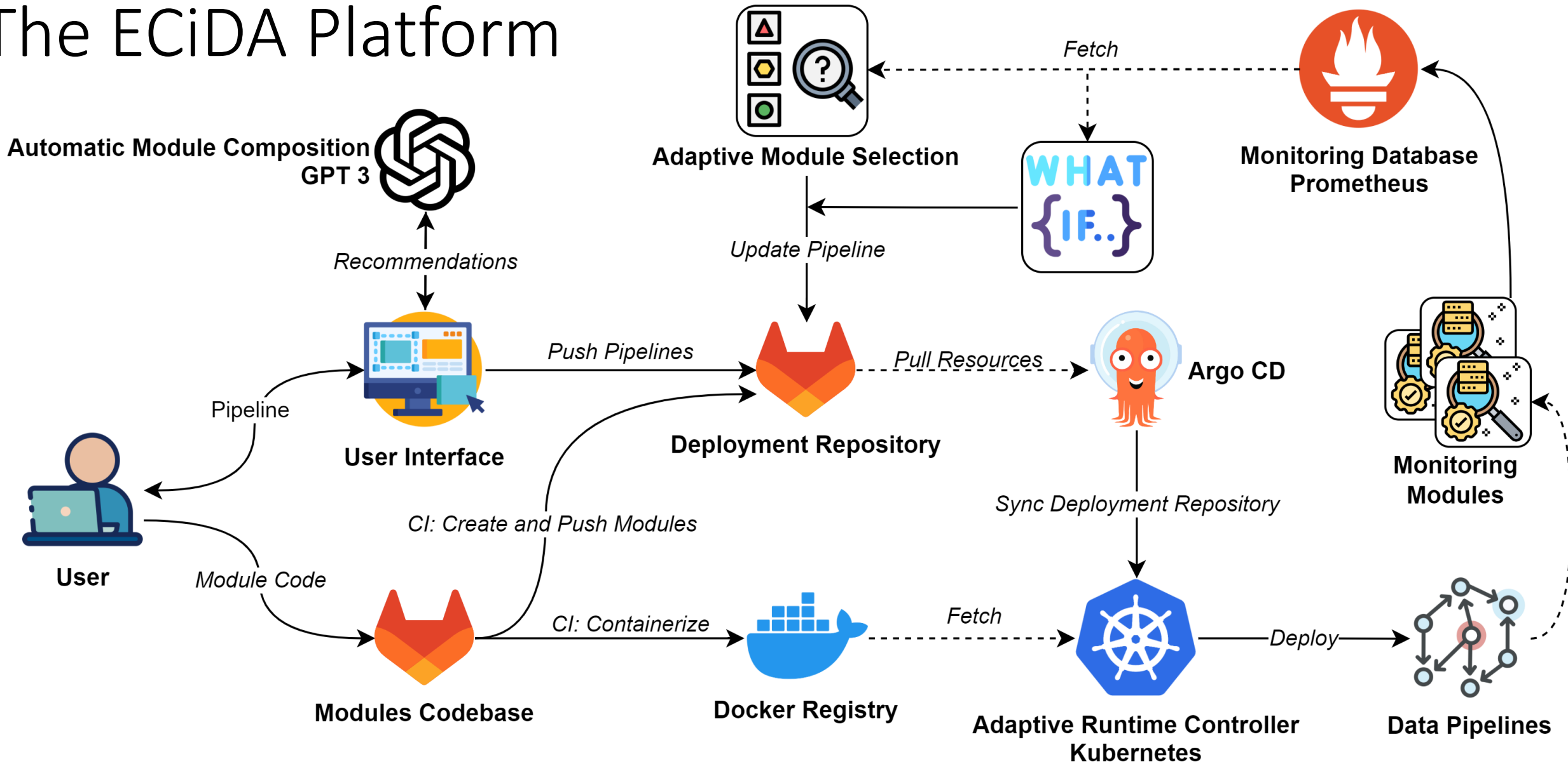


Life Cycle Management





The ECiDA Platform





User Interface

ECiDA > Project Name > Edit Project > consumer-v2 Discard Save

Add modules

Filter Match all

Scaler Monitor env:management Clear

Search for definitions

Implementations (0 selected) Multi-select

- GeneralScaler
- MinMaxScaler **Popover title**
Use the adaptive module selector to find a fitting Scaler
- GeneralMonitor
- Adaptive Scaler
- Adaptive Monitor

Clear Selection Add

Canvas

Pipeline > Main

Suggest

Summary Inputs Outputs +

income	time-series
harmonic_mean	int

new_me|

Get a pipeline suggestion by giving a description of the desired result.

Discard Suggest

Component Data Consumer

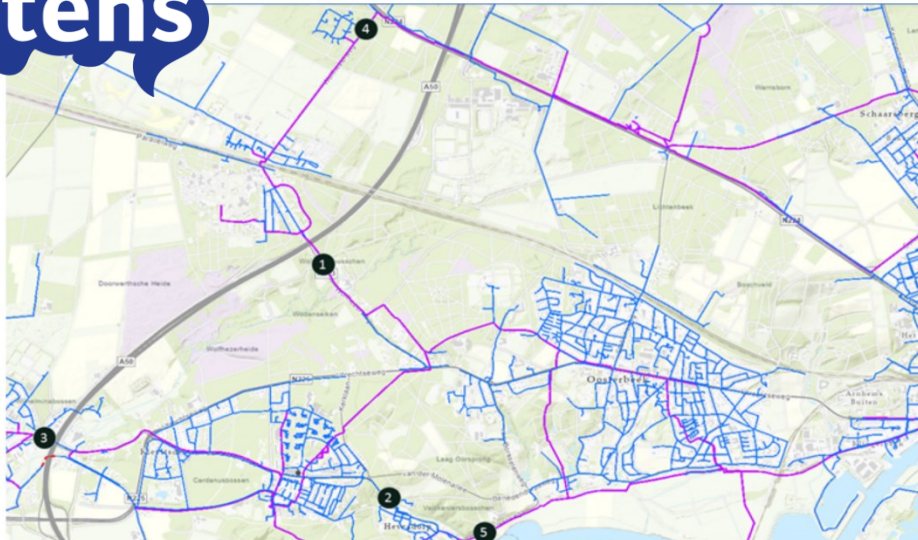
Module: consumer-v2

Standard implementation of a subscriber in the publisher-subscriber pattern. Provides simple printing of any data sent

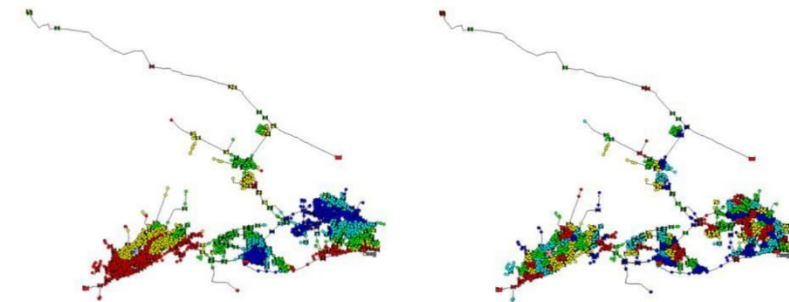
Monitor

Real-World Use case

Pressure Estimation in Water Distribution Networks Using Deep Learning on Graphs by Huy Cuong Truong and Andrés Tello



5855 Node
290 Sensors



Pressure at random scene t1 Pressure at random scene t2




avg. error: -0.1
Histogram of relative error per node

REL_ERROR
1.0
0.8
0.6
0.4
0.2
0.0



Summary

- The Gap exists
- Engineering aspects of data science are overlooked
- There are enough specifics for us to solve it
 - Module Composition
 - Life Cycle Management
 - Training vs Usage
 - Versioning
 - Continuous Monitoring
 - Experiment Tracking
 - What-if Scenarios

A conceptual image showing a person's arm reaching from a laptop screen to shake hands with a person sitting at a wooden table. The laptop is a silver MacBook Air. On the table, there is a smartphone, a pen, and a white coffee cup. The background is a blurred outdoor setting.

University of Groningen
Distributed System Group
Mostafa Hadadian Nejad Yousefi
m.hadadian@rug.nl