Cross-modular Applications of AI and ML in B5G networks

Edwin Yaqub, RapidMiner Research

27 May 2020

Presentation Layout

- Introduction to (EU-H2020) project ARIADNE
  - Vision
  - Use Cases
- Artificial Intelligence (AI)/Machine Learning (ML) Application Areas
  - AI/ML Landscape and Disciplines
  - Predictive Analytics, Prescriptive Analytics and Predictive Optimizations
- Approaching AI and ML
- Cross-Modular Concerns
- Standardization Initiatives
- RapidMiner Data Science Platform
  - Automated ML
  - Management and Orchestration Tools

Introduction to Project ARIADNE

5G PPP

ARIADNE

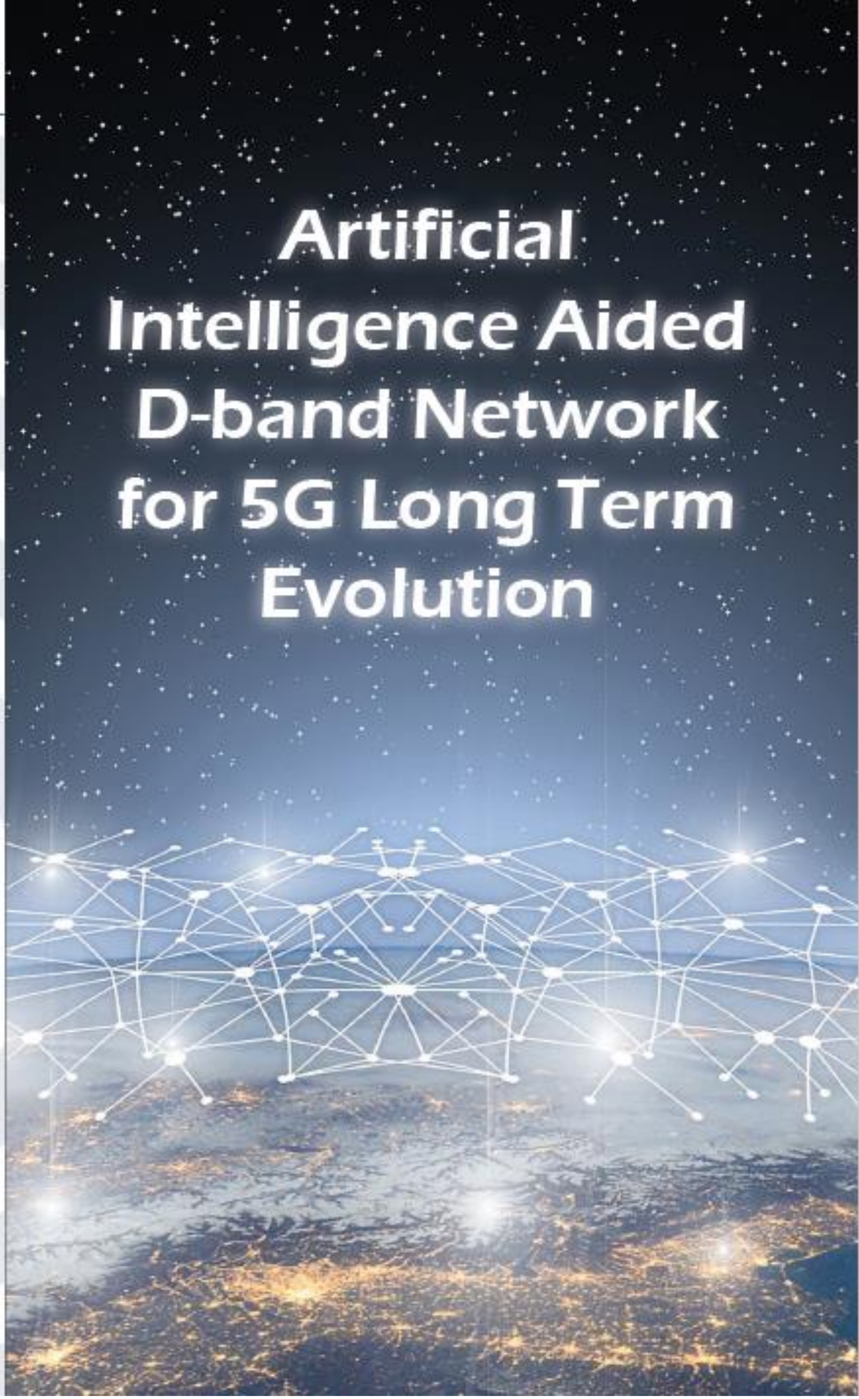
Coordinator
Dr. Halid Hrasnica
Eurescom, Heidelberg, Germany

Scientific and Technical Project Manager
Prof. Dr. Angeliki Alexiou
University of Piraeus Research Centre,
Athens, Greece

Website: https://www.ict-ariadne.eu
Twitter: @AriadneIct
Email: contact@ict-ariadne.eu

Partners

INTRACOM TELECOM
Fraunhofer IAF
cnrs
NOKIA
rapidminer
DEMOKRITOS
Telefónica
Telefónica Investigación y Desarrollo
EURESCOM
Aalto University School of Electrical Engineering
UNIVERSITY OF OULU
POLITÉCNICA

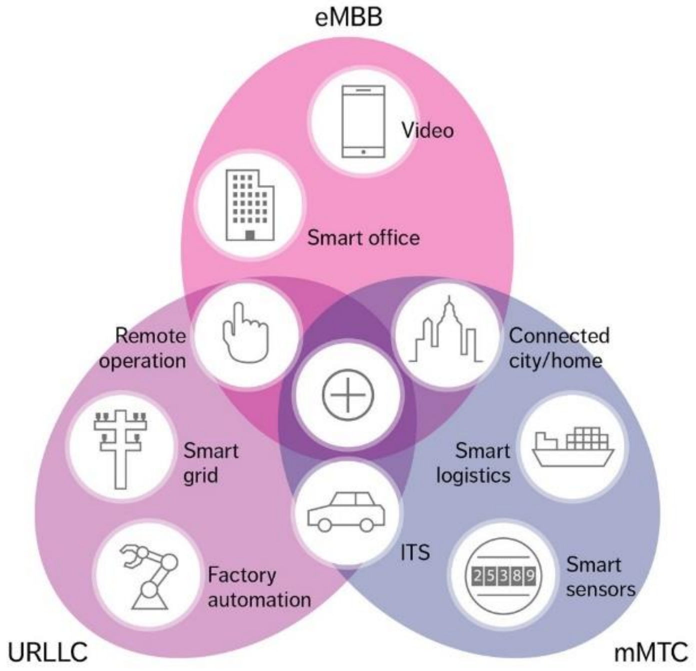


ARIADNE is a three years Research and Innovation action / project under the EU program Horizon 2020 (Grant Agreement no. 871464) started on 1 November 2019

Introduction to ARIADNE project

- Vision

***Ar**tificial **I**ntelligence **A**ided **D**-band **N**etwork for 5G Long Term **E**volution is a H2020 5G PPP project which aims to bring together **a novel high frequency radio architecture**, an advanced wireless **connectivity based on reconfigurable metasurfaces**, and an **enhanced network management supported by AI** to establish a new type of **intelligent communications system beyond 5G**.*



RoadMap: Vision to Objectives and Actions

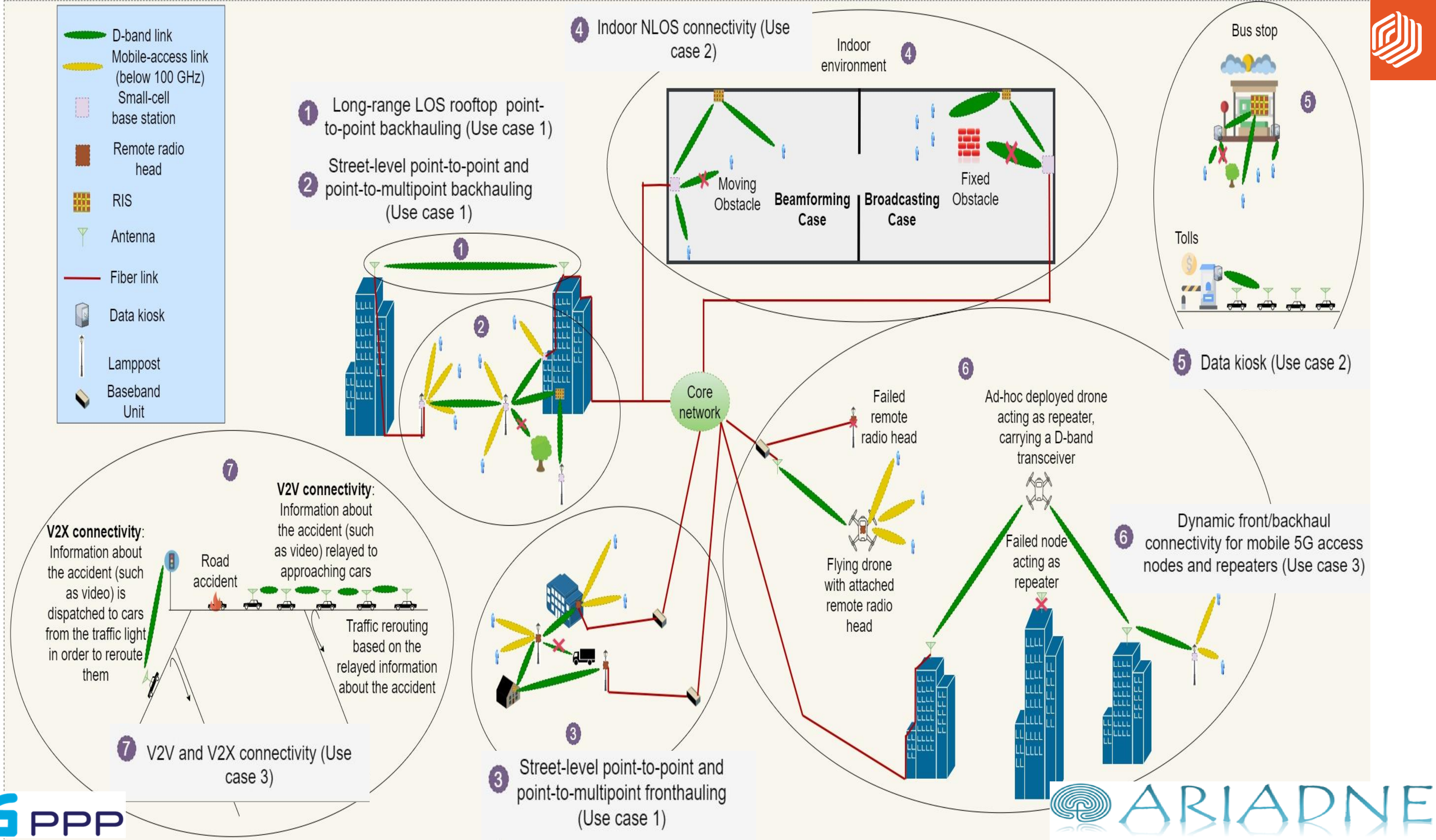
Pillars	Objectives	Actions	Demos	Main KPIs
D-band for 100 Gbit/s reliable wireless connectivity	100 Gbit/s capable, energy and spectral-efficient D-Band wireless B5G networks	• D-band front-end • Baseband DSP • Adaptive spatial SP • High gain D-band antennas • Channel modelling	P2P D-band LOS outdoor demo	100 Gbps throughput 100m range Massive number of devices Zero latency Ultra-high reliability
Communications beyond the Shannon paradigm	Ultra-reliable D-band connectivity and reconfigurability in all usage environments	• Reconfigurable adaptive metasurface design • Beamforming for LOS and NLOS links • MAC design	Metasurface-based D-band demo	
Artificial Intelligence-based wireless system concept	Transform networks B5G into intelligent connectivity/computing platforms	• ML for channel modelling • ML-based resource allocation and energy efficiency ML for network deployment optimization	Intelligent D-band Network Demo	

Use Cases

- **Use case 1**: Outdoor backhaul/fronthaul networks of fixed topology
  - Scenario 1: Long-range LOS rooftop point-to-point backhauling.
  - Scenario 2: Street-level point-to-point and point-to-multipoint backhauling/fronthauling.

- **Use case 2**: Advanced NLOS connectivity based on metasurfaces
  - Scenario 1: Indoor advanced NLOS connectivity based on metasurfaces
  - Scenario 2: Data kiosk

- **Use case 3**: Adhoc connectivity in moving network topology
  - Scenario 1: Dynamic front/backhaul connectivity for mobile 5G access nodes and repeaters
  - Scenario 2: V2V and V2X connectivity


D-band link
Mobile-access link (below 100 GHz)
Small-cell base station
Remote radio head
RIS
Antenna
Fiber link
Data kiosk
Lamppost
Baseband Unit
1 Long-range LOS rooftop point-to-point backhauling (Use case 1)
2 Street-level point-to-point and point-to-multipoint backhauling (Use case 1)
4 Indoor NLOS connectivity (Use case 2)
Indoor environment
Moving Obstacle
Beamforming Case
Broadcasting Case
Fixed Obstacle
Bus stop
Tolls
5 Data kiosk (Use case 2)
Core network
Failed remote radio head
Ad-hoc deployed drone acting as repeater, carrying a D-band transceiver
Flying drone with attached remote radio head
Failed node acting as repeater
6 Dynamic front/backhaul connectivity for mobile 5G access nodes and repeaters (Use case 3)
V2X connectivity: Information about the accident (such as video) is dispatched to cars from the traffic light in order to reroute them
Road accident
V2V connectivity: Information about the accident (such as video) relayed to approaching cars
Traffic rerouting based on the relayed information about the accident
7 V2V and V2X connectivity (Use case 3)
3 Street-level point-to-point and point-to-multipoint fronthauling (Use case 1)
5G PPP
ARIADNE

AI/ML Application Areas

AI/ML Application Areas

- Where can we apply AI/ML?
  - Channel modeling
    - Estimating parameters of the channels
    - Profiling adverse effect of weather on channels
  - Beamforming (assigning beams to users) and ray tracing (follow a mobile node)
  - Modeling behavior of RIS (Reconfigurable Intelligent Surface) that uses Metasurface.
  - Network optimization
    - Resource allocation or Route finding/scheduling using performance criteria (e.g. energy saving, reliability)
    - Placement of radio network components (e.g. to maximize signal strength at key operational areas)
    - Performing offline and online optimizations (for dynamic cases)
    - Optimizing with and without RIS (for indoor beamforming cases).

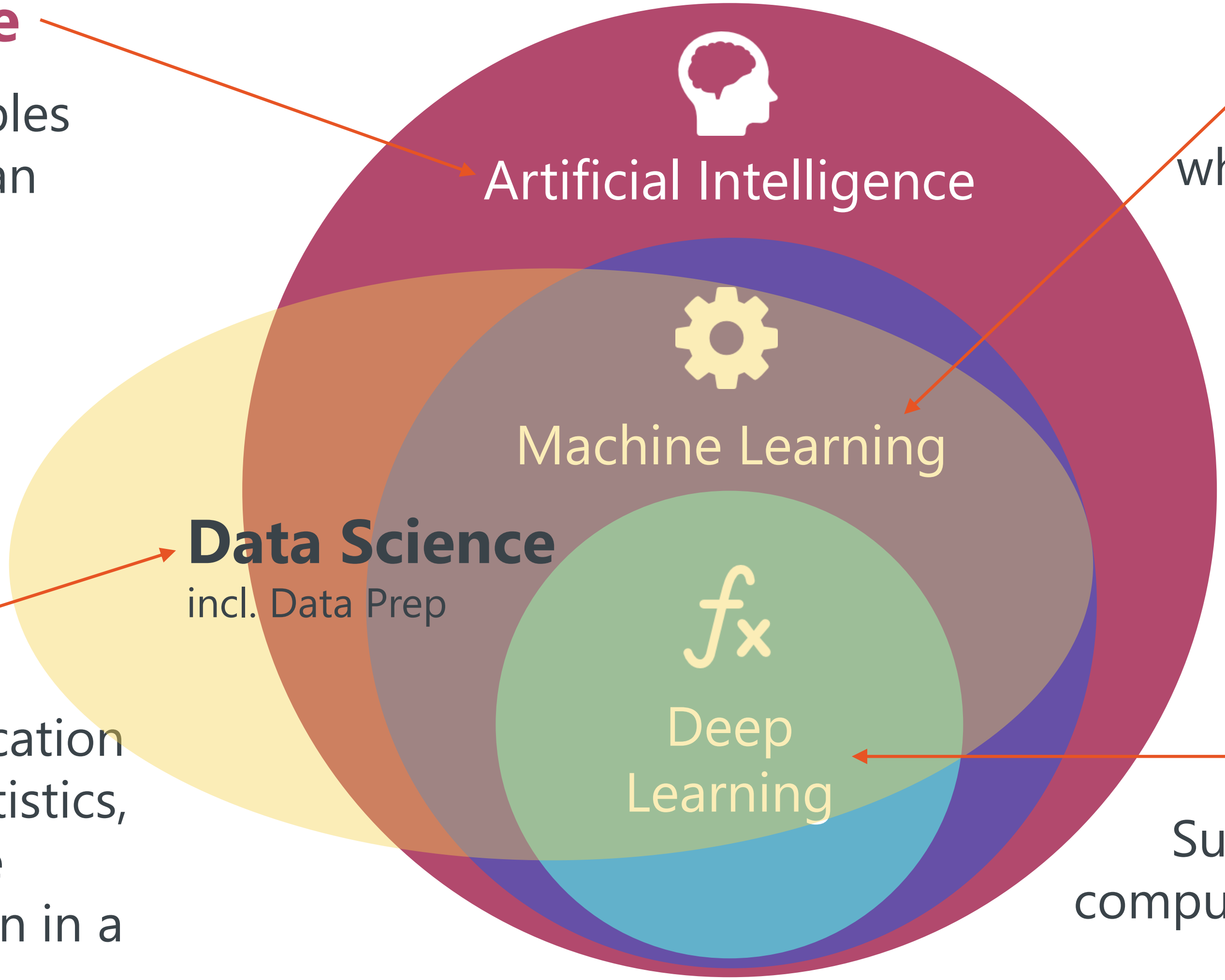
AI/ML Landscape

Artificial Intelligence

Any technique which enables computers to mimic human behavior.

Machine Learning

Subset of AI techniques which use statistical methods to enable machines to improve with experiences.



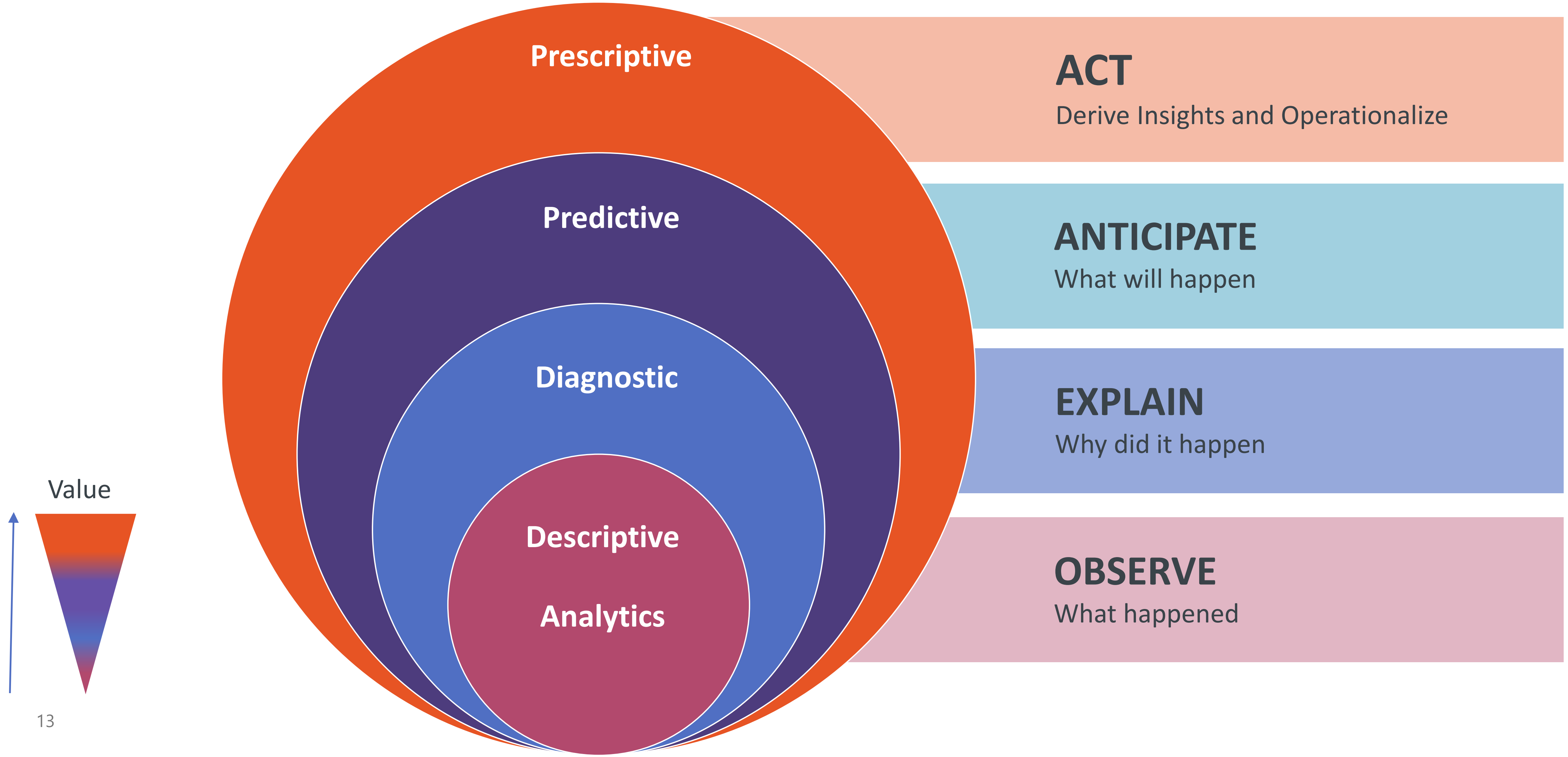
Data Science

Covers the practical application of advanced analytics, statistics, machine learning, and the necessary data preparation in a business context.

Deep Learning

Subset of ML which make the computation of multi-layer neural networks feasible.

AI/ML Disciplines



AI/ML Application Areas (revisited)

- **Predictive Analytics**
  - Forecasting
    - Demand for connections or data transmission rates at different time periods
    - Demand for energy consumption in the network
  - Predicting
    - Bottlenecks at different nodes and congestion in the network
    - Properties of a link: blockages, reliability, failure rate, service degradation (e.g., in terms of packet loss, effective radius, etc.)
    - Estimated parameters of a multi-path channel.
    - Attenuation level in signal quality due to weather affects.
    - Machine failures before they happen, to proactively prevent failure and save repair costs.
    - Movement (e.g. direction, angle) of user or mobile node
  - Detecting
    - Anomalous traffic flows

AI/ML Application Areas (revisited)

- **Prescriptive Analytics**
  - What-If Analysis to derive insights
    - *Interact* with a predictive model to *understand* behavior of complex systems (such as a RIS or Metasurface)
    - *Exploit* predictive model by optimizing predictions for desired outcomes by injecting business constraints
    - Get optimal inputs (generate recipe)

- **Predictive Optimizations**
  - Resource allocation / Route scheduling
    - Include predictions from ML models within fitness function to get superior solutions
    - Include RIS as part of network
  - Dynamic environments: Real-time planning and optimization
    - Stochastic and non-stochastic variants.

Approaching AI and ML (Informal Tips and Standard Method)

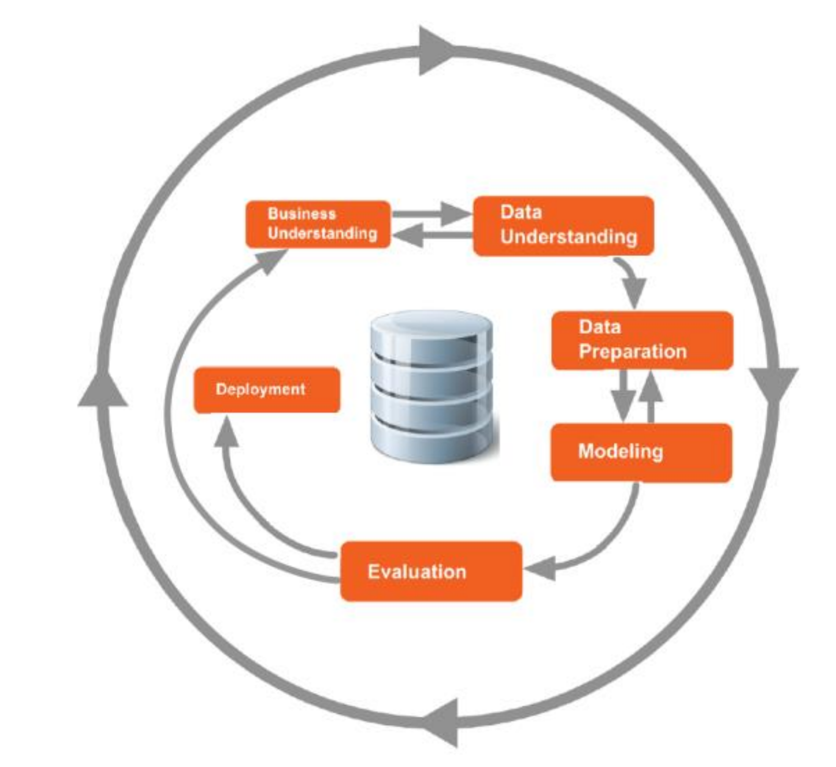
Approaching AI and ML

- Informal tips on how to *identify* AI/ML problems
  - Problems that have not been solved so far in the literature
  - Attempt to solve problems that are currently solved inefficiently
  - Objective: Solve problems efficiently and smartly

- Formal methodology to *describe* and *solve* analytics problems, especially in inter-disciplinary teams
  - **CRISP-DM: CR**oss-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining
  - https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

CRoss-Industry Standard Process for Data Mining

- **CRISP-DM** is a standardized approach to tackle data mining (analytics) problems
  - Leading method used by data mining industry
  - A conceptual and cyclic method independent of data or tools



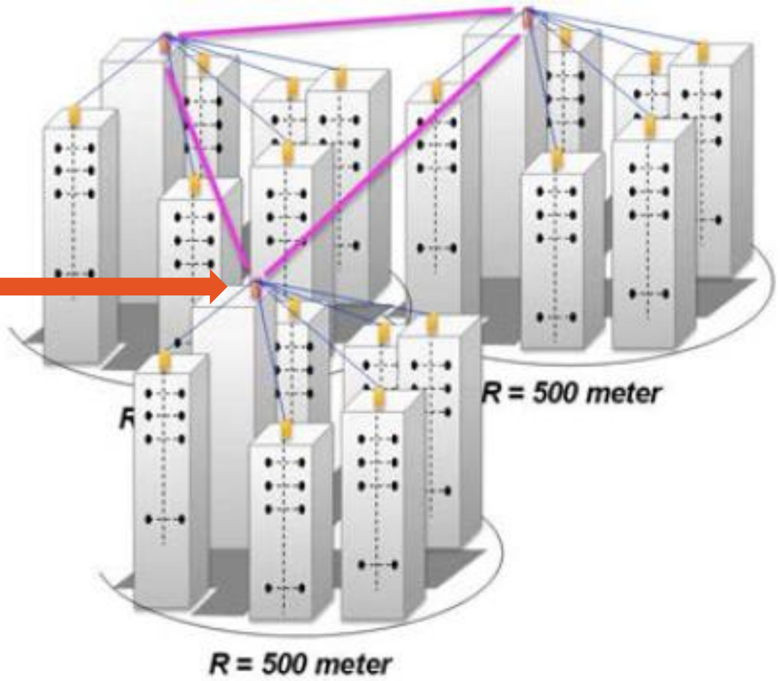
- Relationship between steps can be iterative
- Start with Business Understanding
- Normal to go back to Business Understanding before Deployment
- Experience from Deployment triggers revisions
- Subsequent solutions improve last ones.

Cross-Modular Concerns

AI/ML Operationalization

- Management and Orchestration aspects of AI/ML models to be considered from start
  - Where will the models be trained, updated, deployed, and monitored?

Cloud could be used for data storage, training and application of ML models – based on scenario requirements

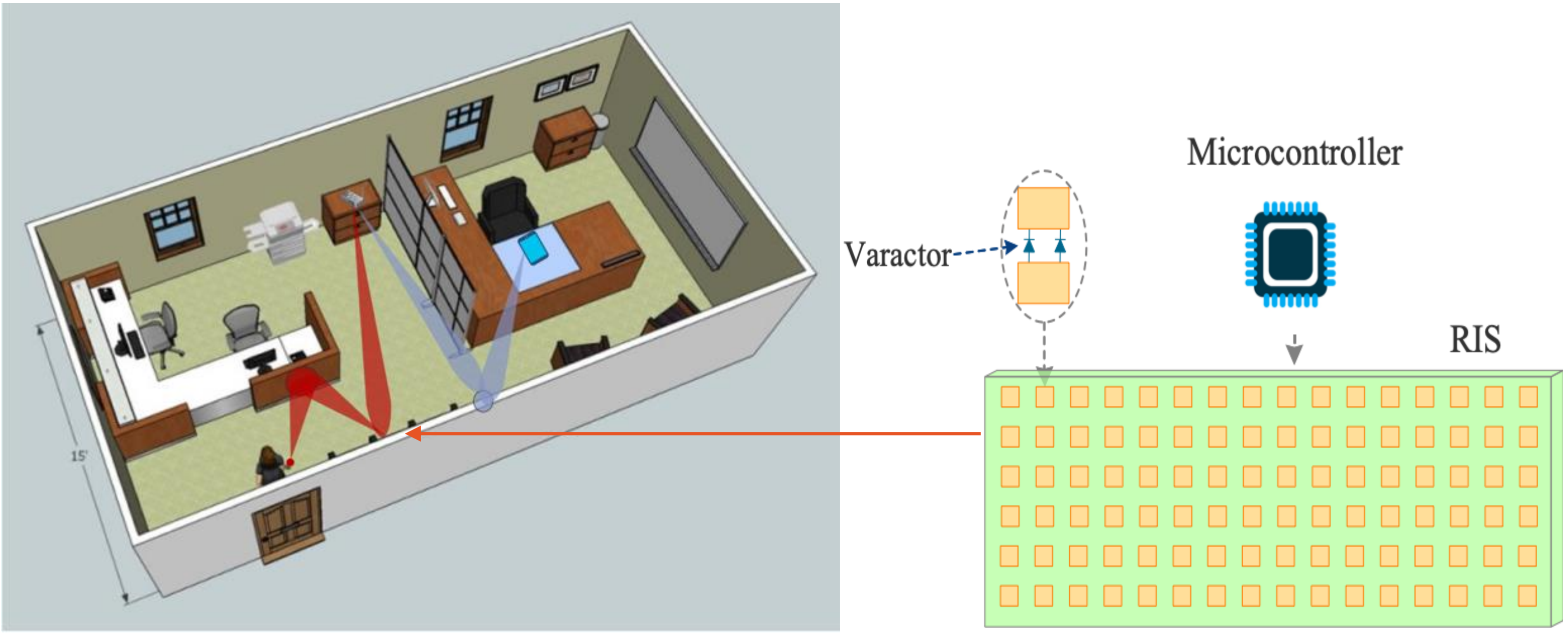


Network topology with nodes at:

- **Core** (more resources to train, update and apply AI/ML models)
- **Edge** (limited resources locally)
- **Cloud** (store data, train, update or apply AI/ML models)

AI/ML Operationalization

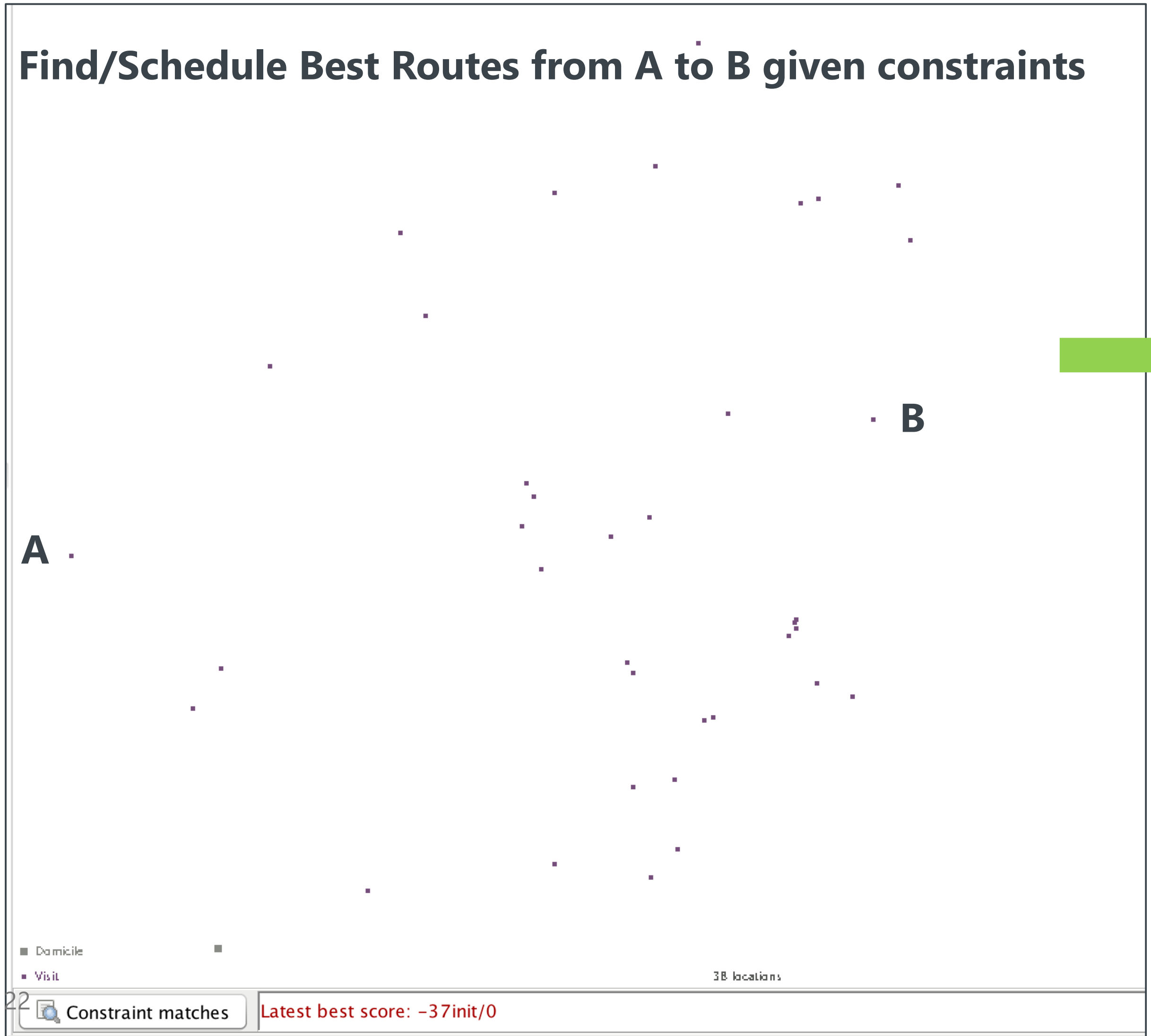
- Deployment aspects of AI / ML models to be considered from start

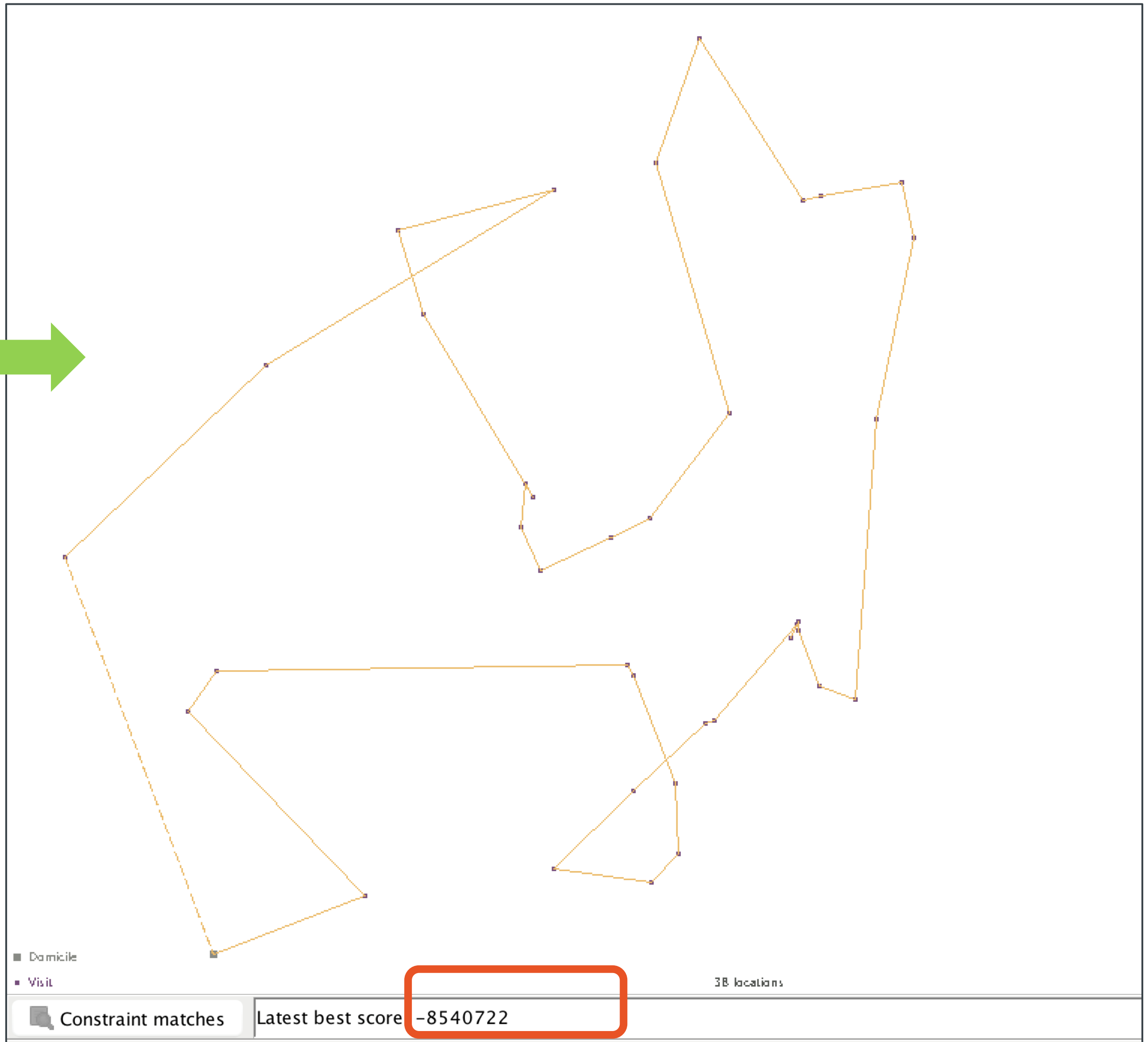


Non-functional Properties:

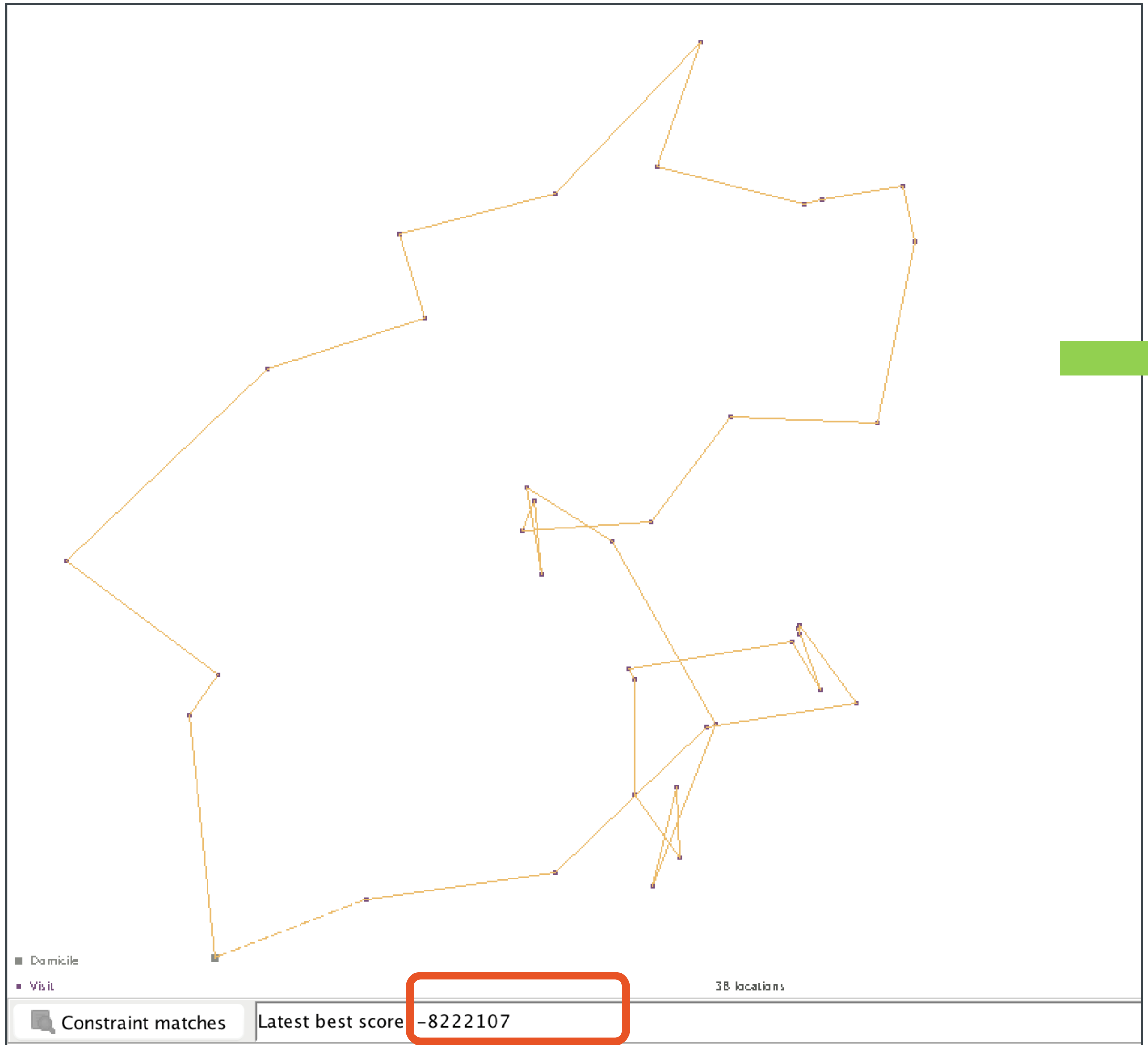
- Prediction / Prescription to be applied very fast at access point or at Microcontroller on chip to keep latencies very low
- 2^96 possible combinations to turn unit-cells ON or OFF (without pruning)
- Decisions to be made in microseconds

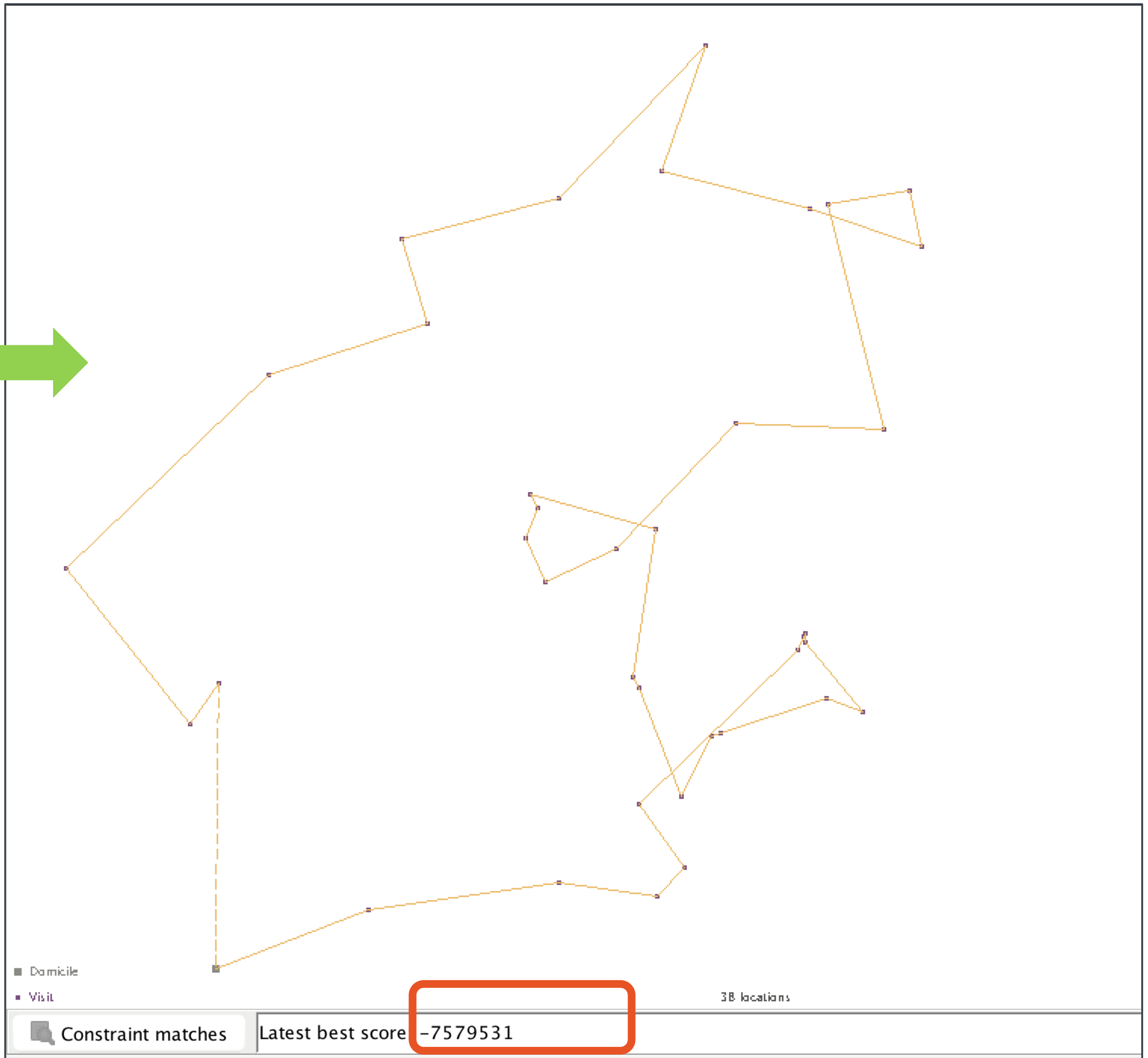
From Classical to AI/ML based Optimization (1)



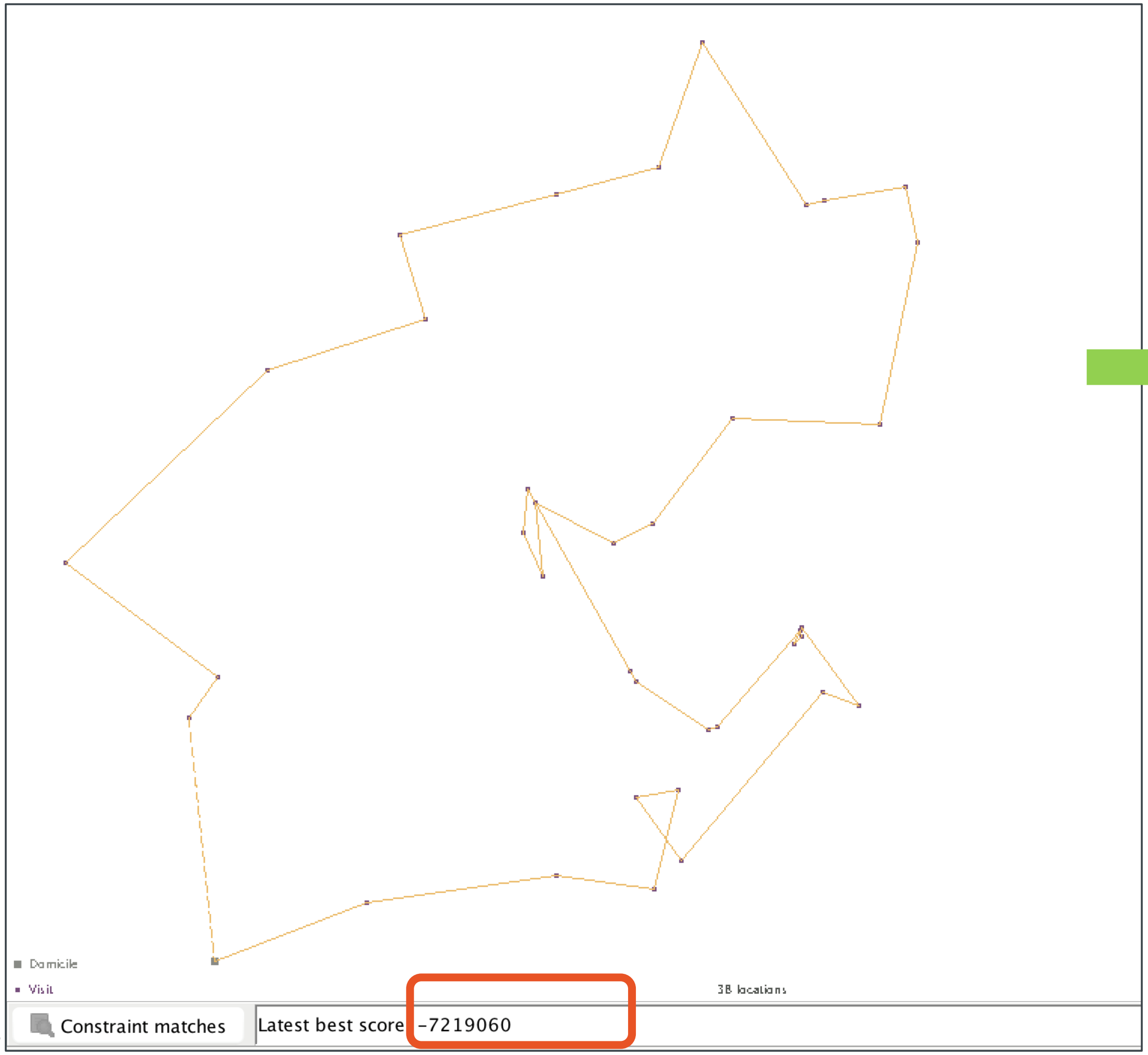


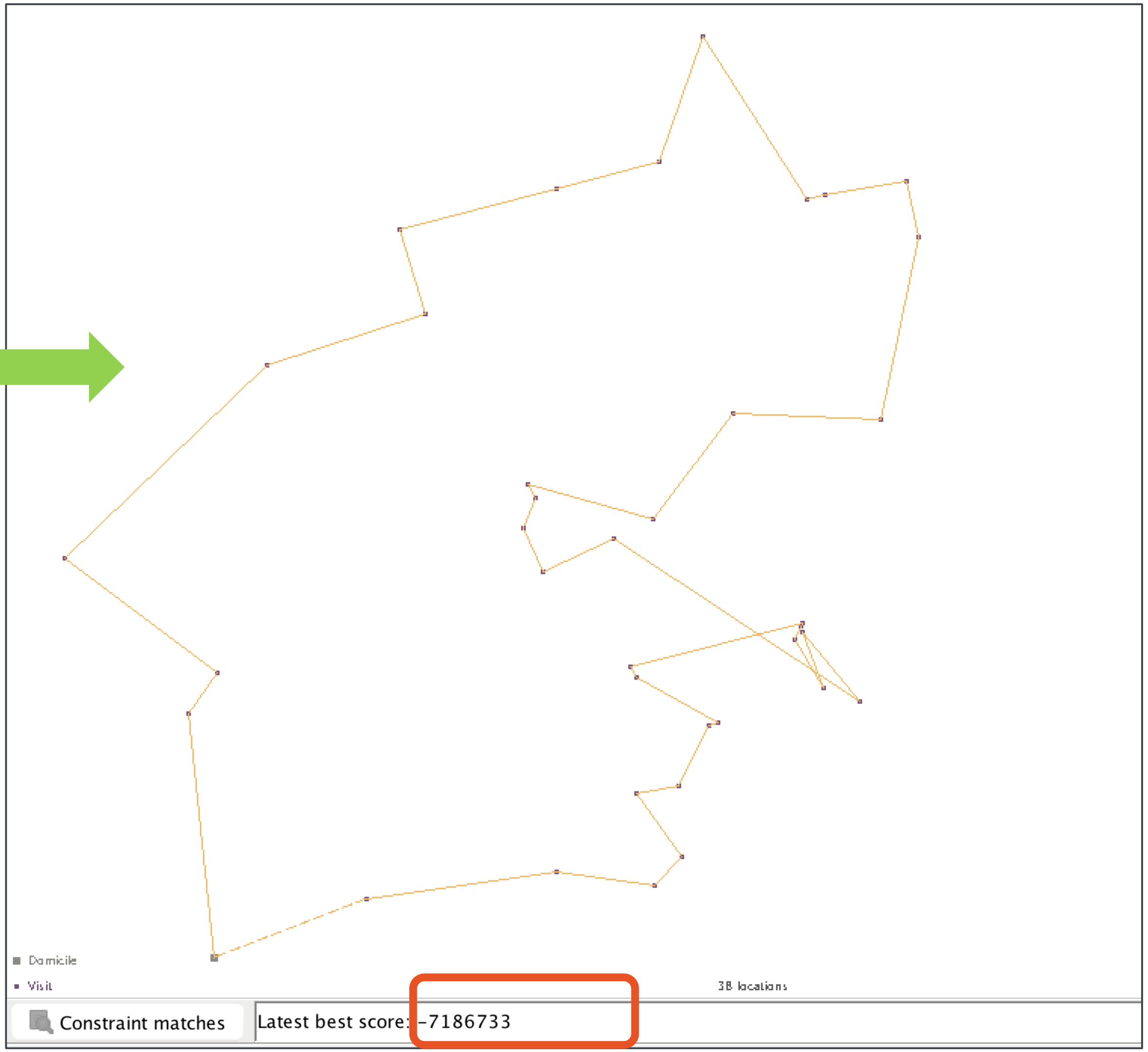
From Classical to AI/ML based Optimization (1)



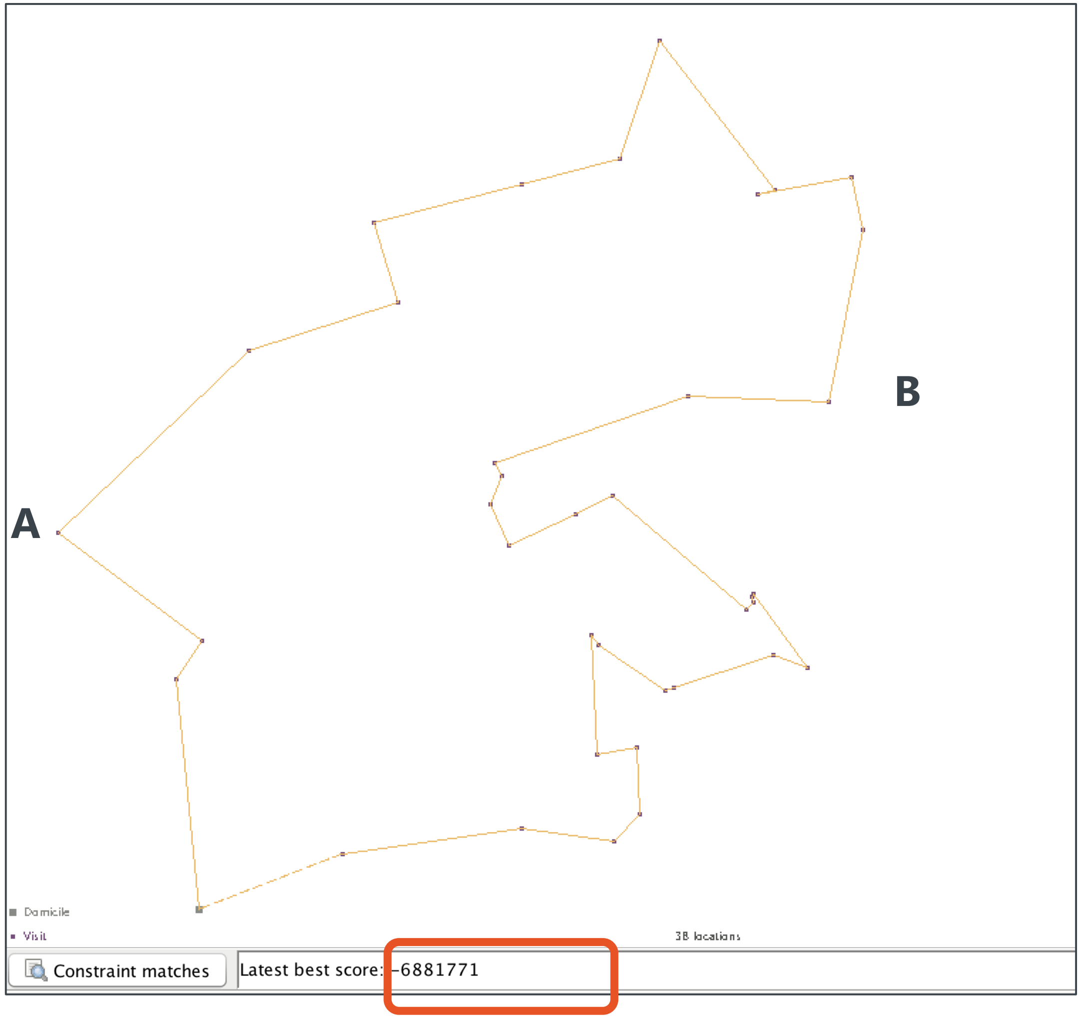


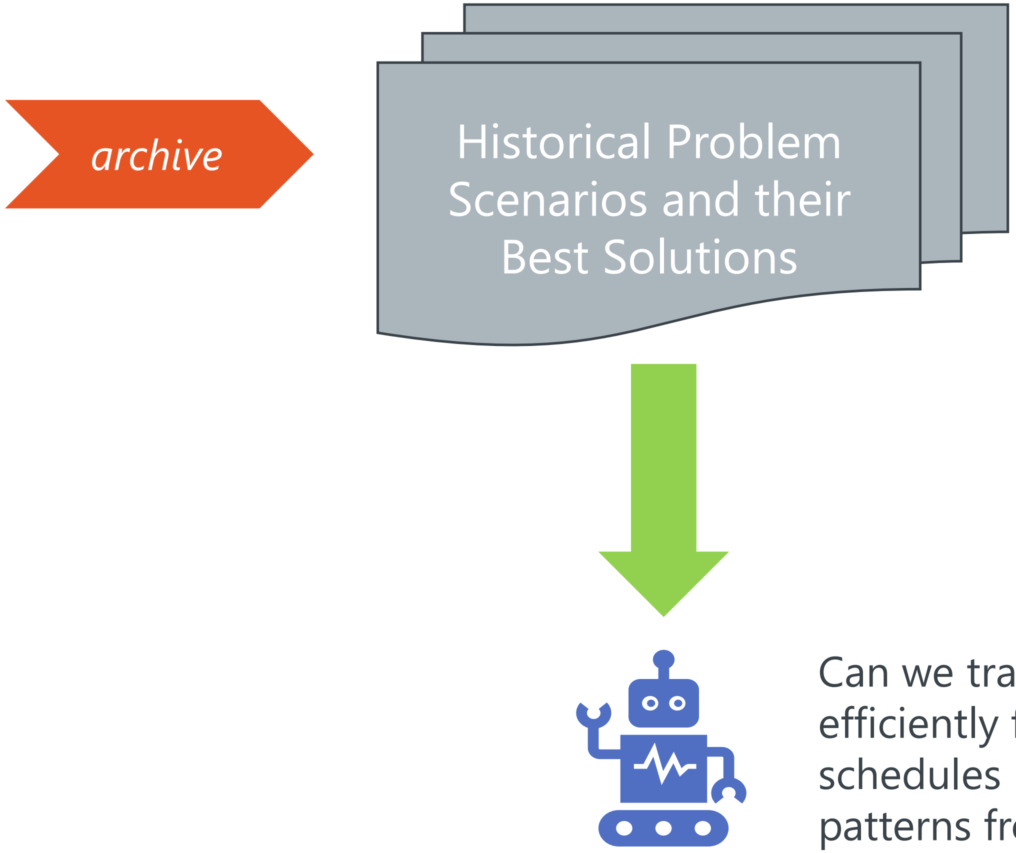
From Classical to AI/ML based Optimization (1)





From Classical to AI/ML based Optimization (1)

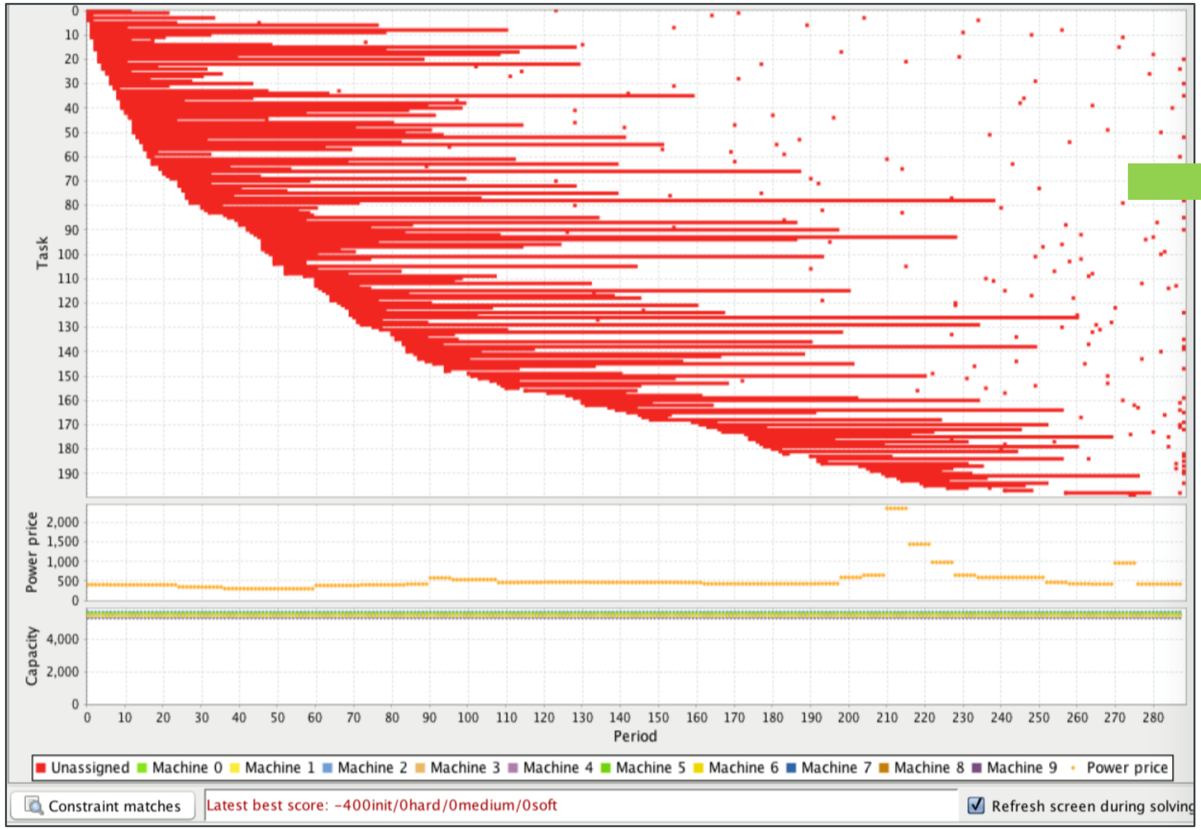




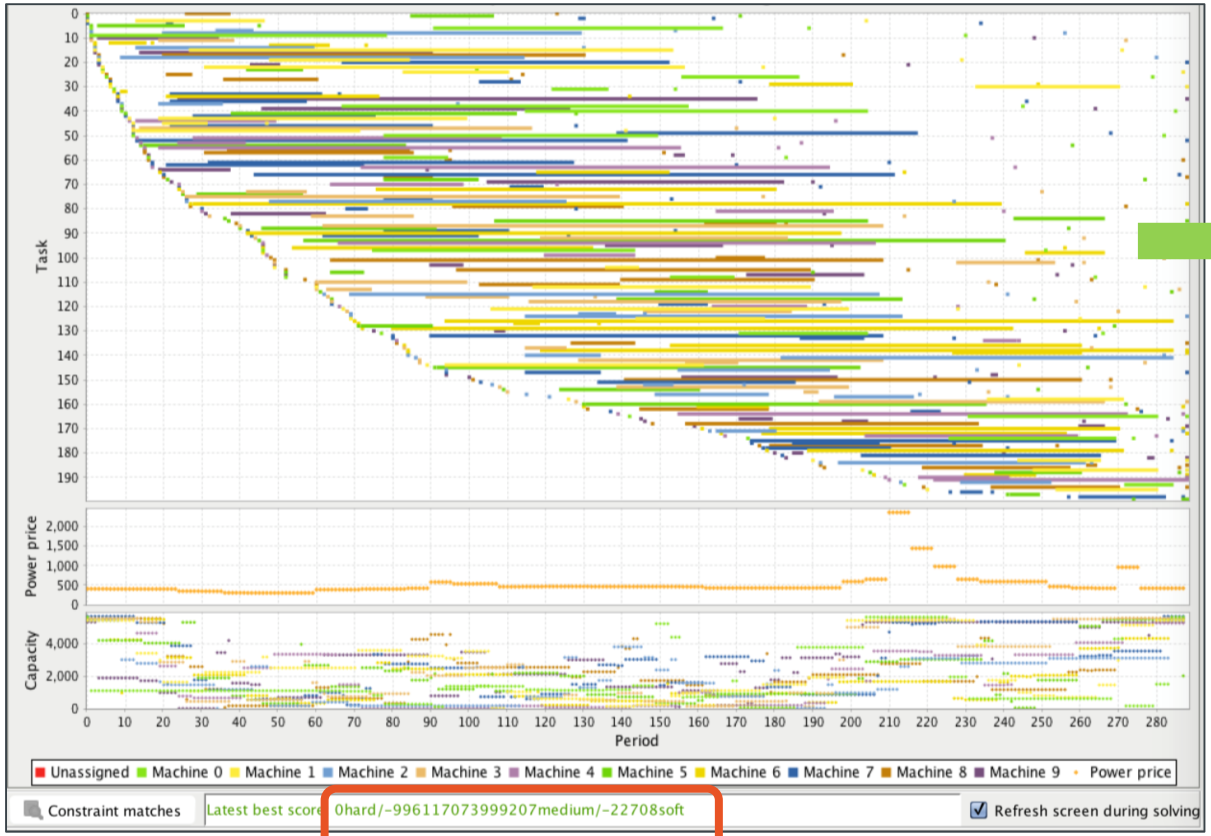
Can we train AI/ML to efficiently find smart schedules by learning patterns from past data?

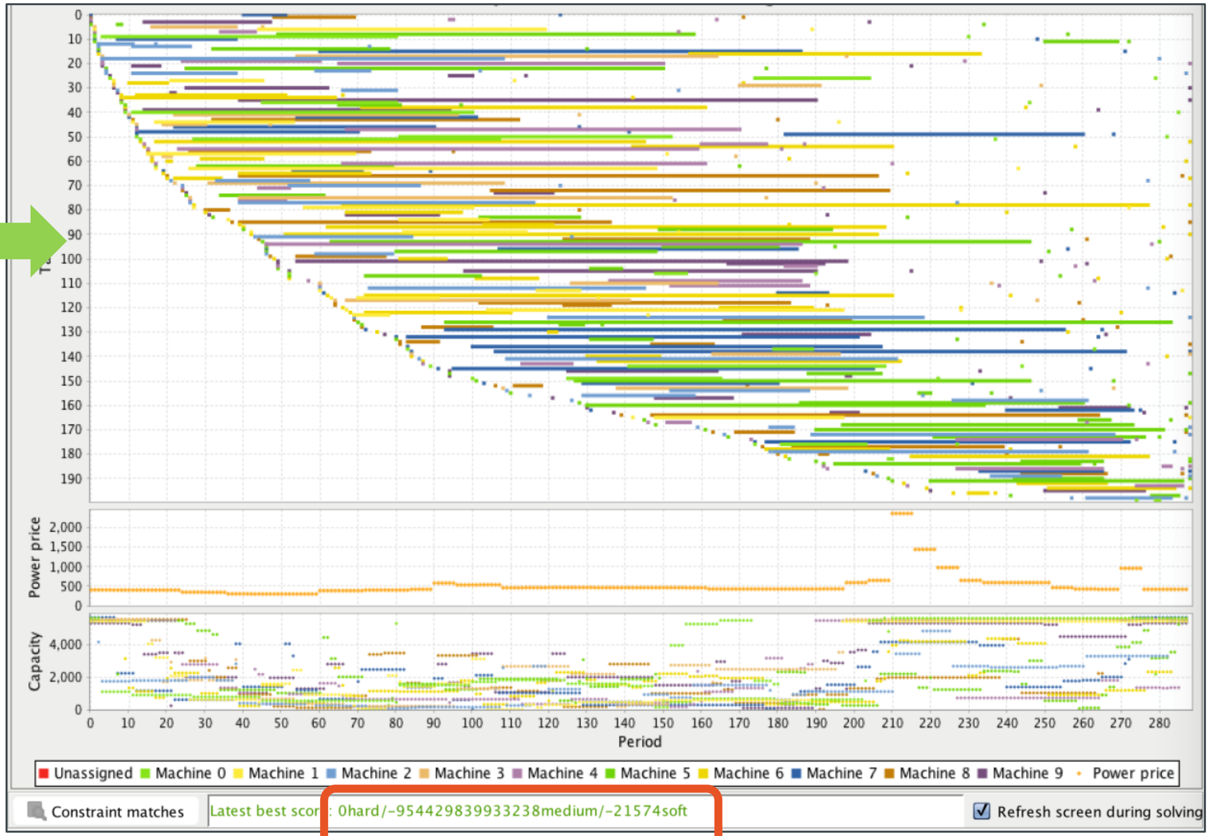
From Classical to AI/ML based Optimization (2)

Allocate tasks to machines over time and reduce power price

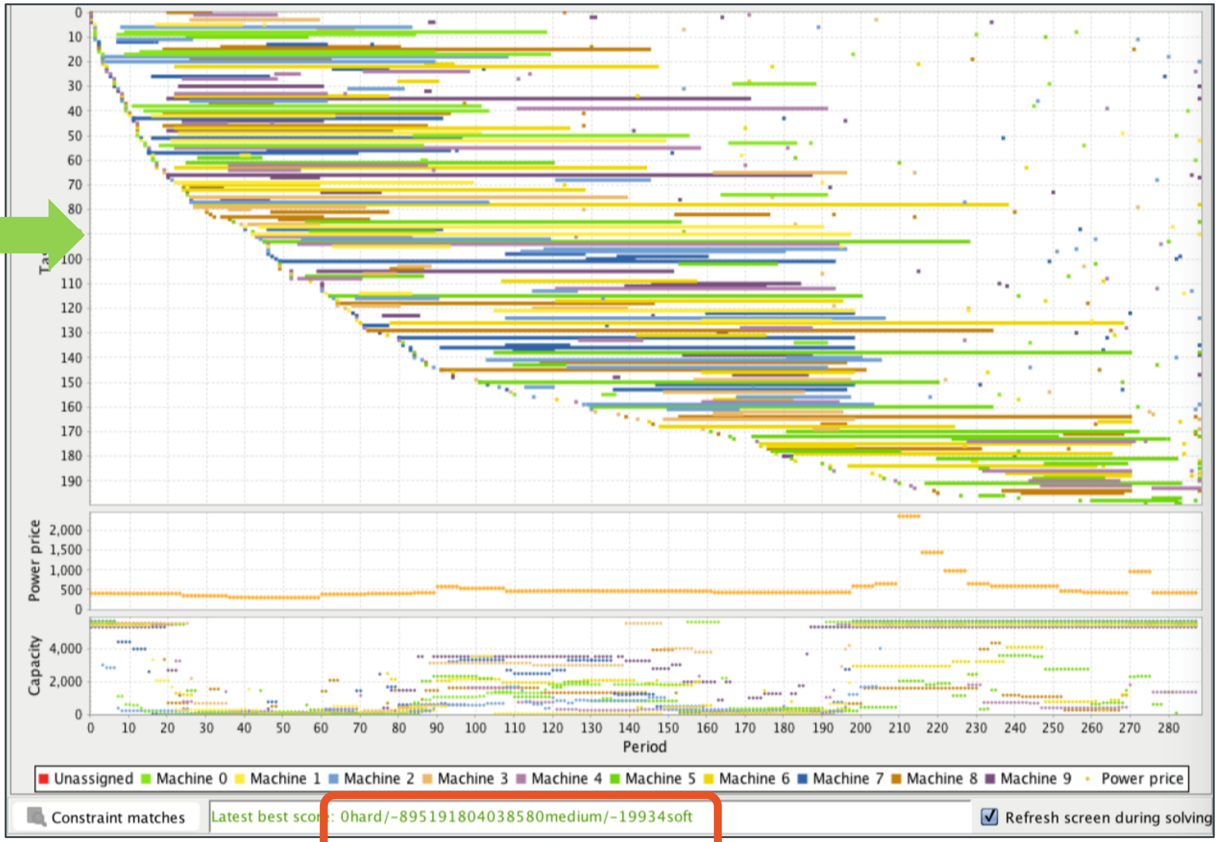


From Classical to AI/ML based Optimization (2)

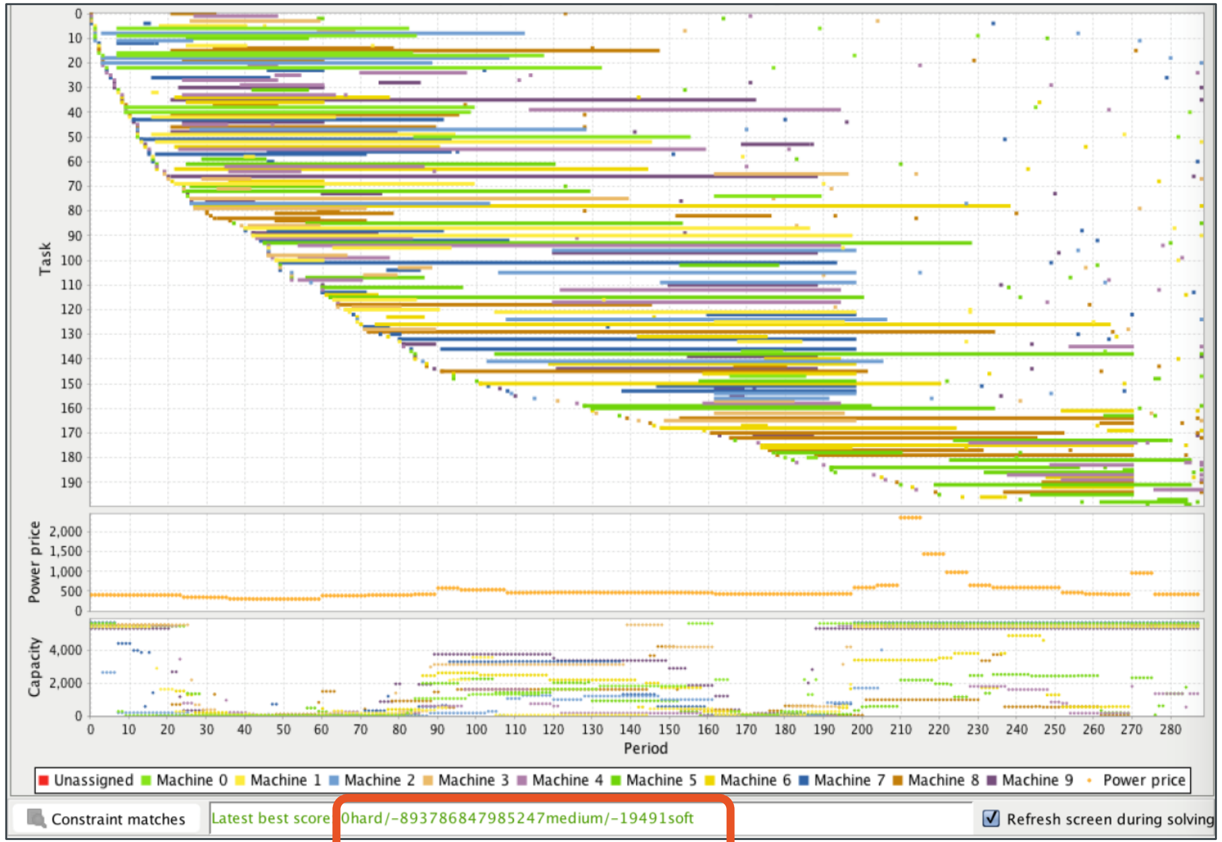


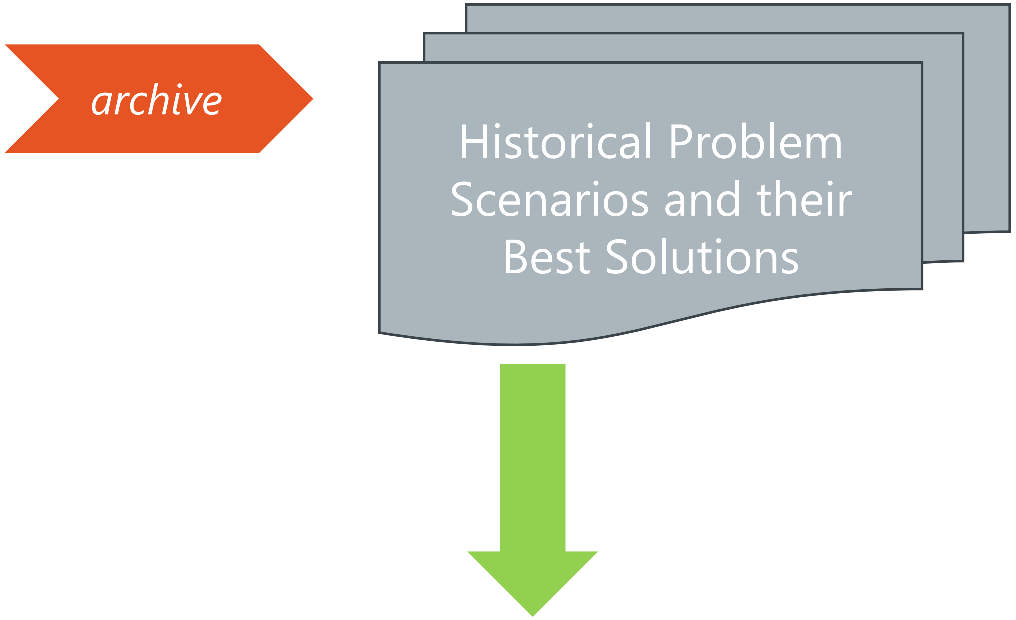


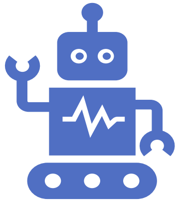
From Classical to AI/ML based Optimization (2)



From Classical to AI/ML based Optimization (2)







Again: Can we train AI/ML to efficiently and smartly allocate resources by learning patterns from past data?

Further R&D Activities in ARIADNE

- Investigating
  - Deep Neural Networks and Auto Encoders
  - Reinforcement Learning
  - Deep Q Learning
  - RNN, Feedforward neural networks,
  - GANS (synthetic data generation close to real systems but anonymous)

Standardization Initiatives

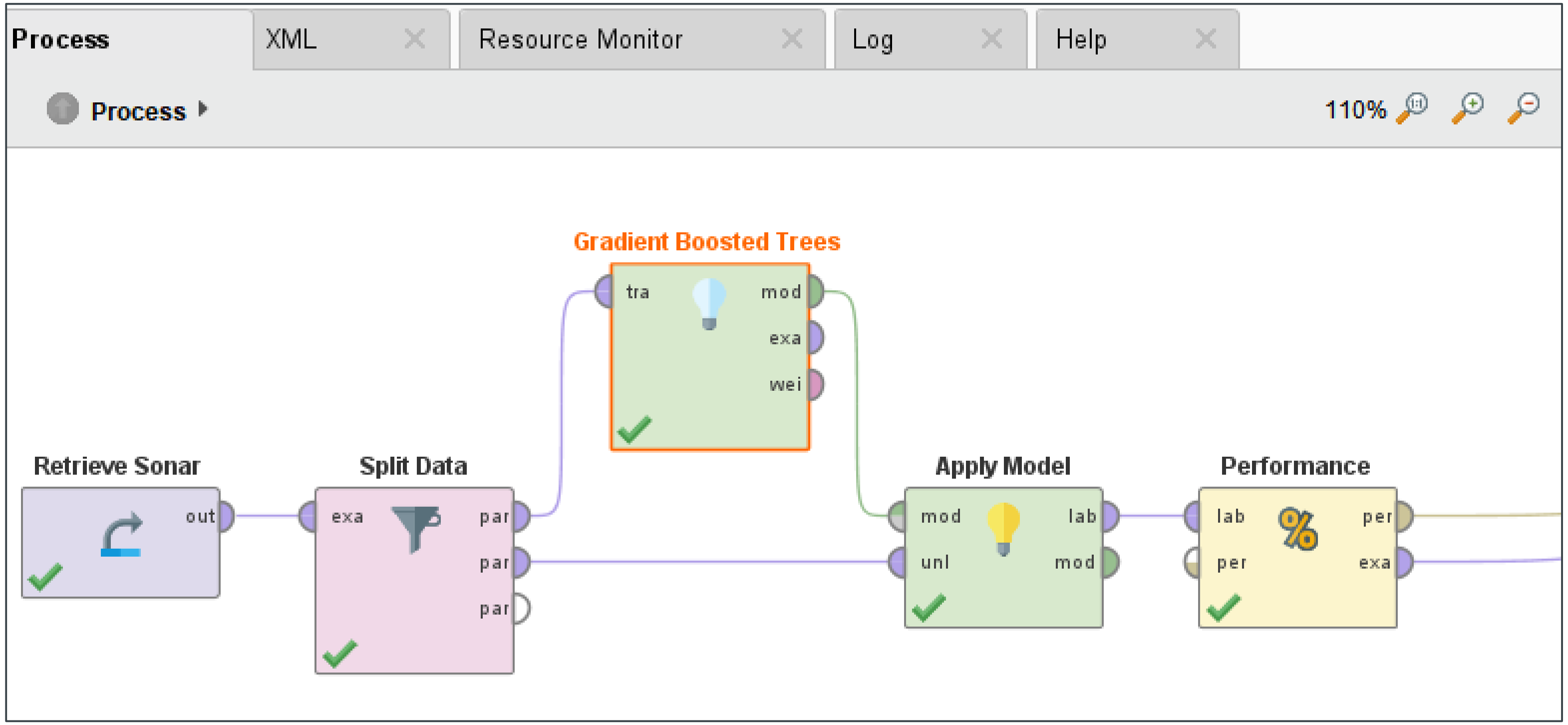
Standardization Initiatives to the Rescue!

- Standardization initiatives (ETSI, ENI, ITU-T) provide frame of reference
  - Focus Group on Machine Learning for Future Networks including 5G
  - *"Architectural framework for machine learning in future networks including IMT-2020"*
    - https://www.itu.int/rec/T-REC-Y.3172/en
  - Highlights challenges and requirements and architectural components. Guides towards standard methods to integrate ML functionalities in future networks!
  - Key Ideas
    - Pipeline-based ML functions, ideally with declarative specifications
    - Support data retrieval and storage from various sources and sinks and ability to plugin support of more sources/sinks
    - ML functionalities to be addressed across multiple network layers/levels, standard methods to train and update models
    - Loose coupling of ML functions with network functions
    - Placement of ML functions at the core, edge or cloud
    - Orchestration of ML functions, including deployment and scaling to keep latencies low
    - Deployment time monitoring of ML models and seamless upgrades (blue-green or primary-secondary deployments)

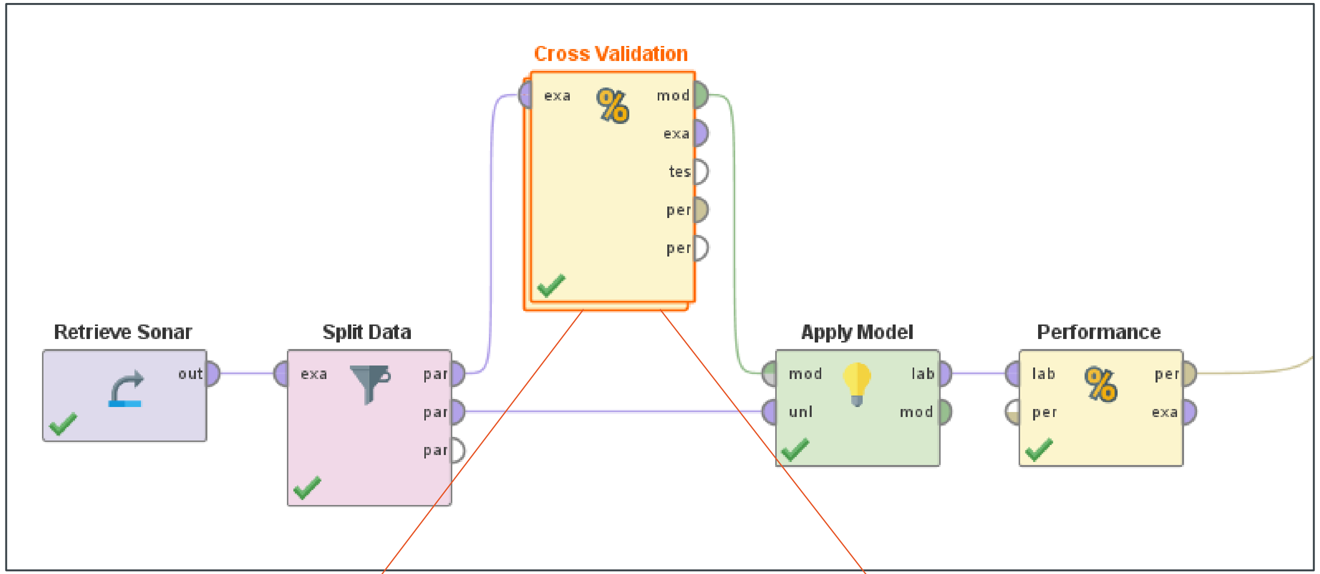
RapidMiner Data Science Platform

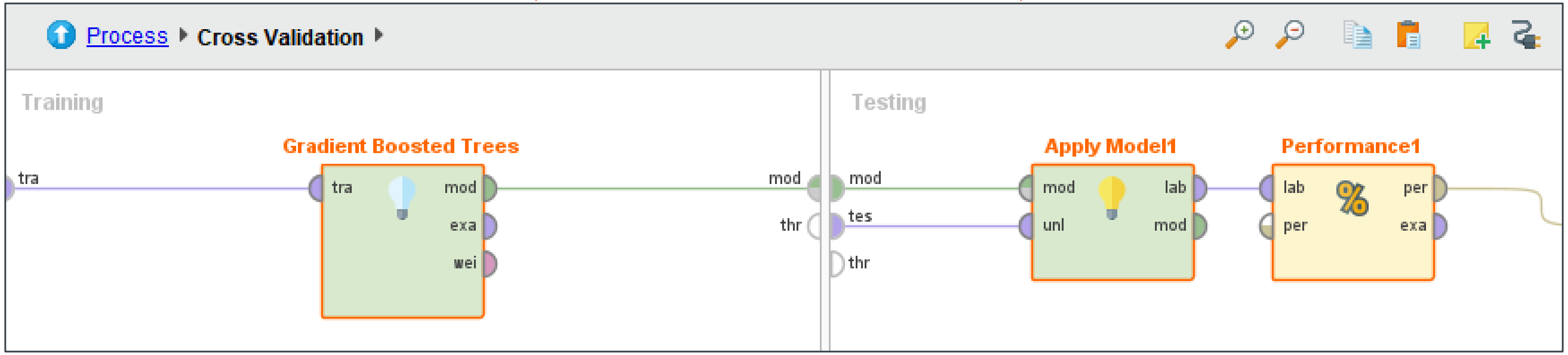
RapidMiner Data Science Platform

- **RapidMiner Studio**: Pipeline based approach with orchestration ecosystem around!
  - Visual building blocks to create AI/ML pipelines (coding-optional, highly modular)
  - Deploy pipelines as webservices in distributed and scalable execution units at Core, Edge or Cloud

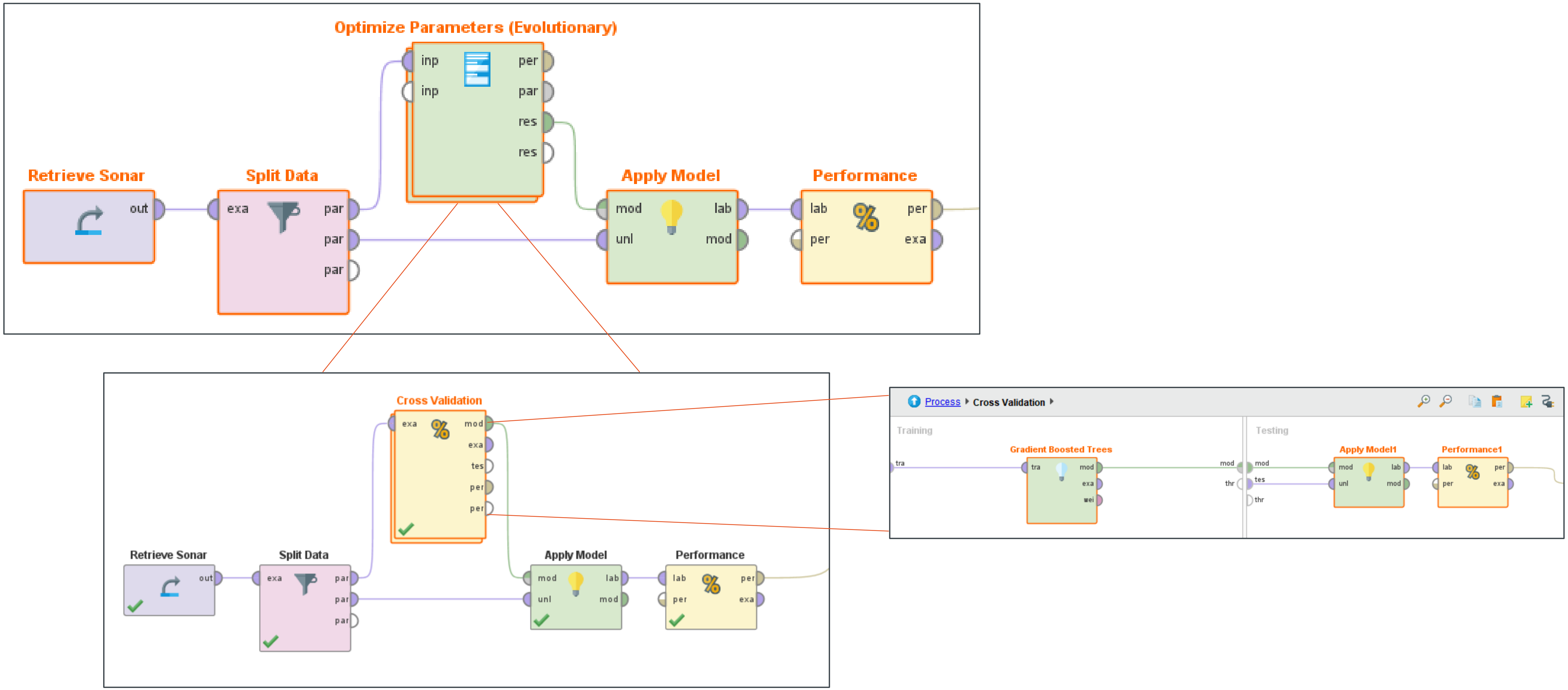


RapidMiner Data Science Platform



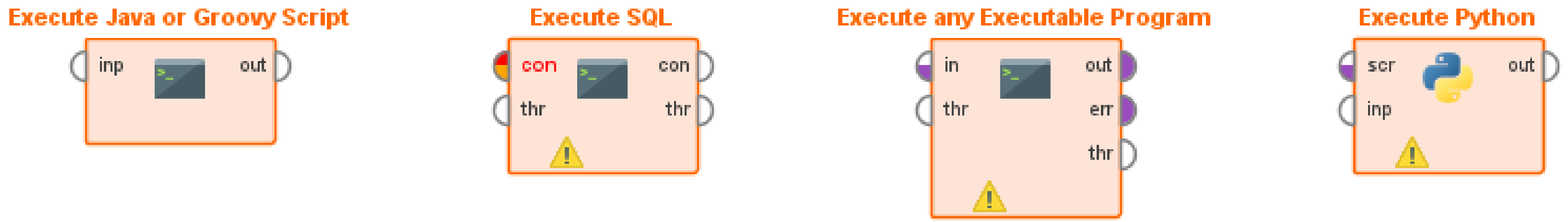


RapidMiner Data Science Platform



Inclusive, Extensible and Pluggable Approach

- Integrate code from popular languages/frameworks/libraries or programs



- Process can be deployed as REST Webservice for integration with components

Automated Machine Learning

Automated Machine Learning

- AutoModel – RapidMiner Tool that democratizes ML
  - Interdisciplinary teams need enablement and guided tooling to attempt AI/ML
  - AutoModel *quickly generates correct ML models* and lets you evaluate them in dashboard
  - No Black-Boxing: Export pipelines or models to Design perspective and tweak it yourself
  - Lets have a short walk-through over few steps to show how easy it is!

File Edit Process View Connections Testing Settings Extensions Help

Views: Design | Results | Turbo Prep | **Auto Model** | Deployments | Hadoop Data

Find data, operators...etc | All Studio

Auto Model

Load Data — **Select Task** — Prepare Target — Select Inputs — Model Types — Results

RESTART | BACK | NEXT

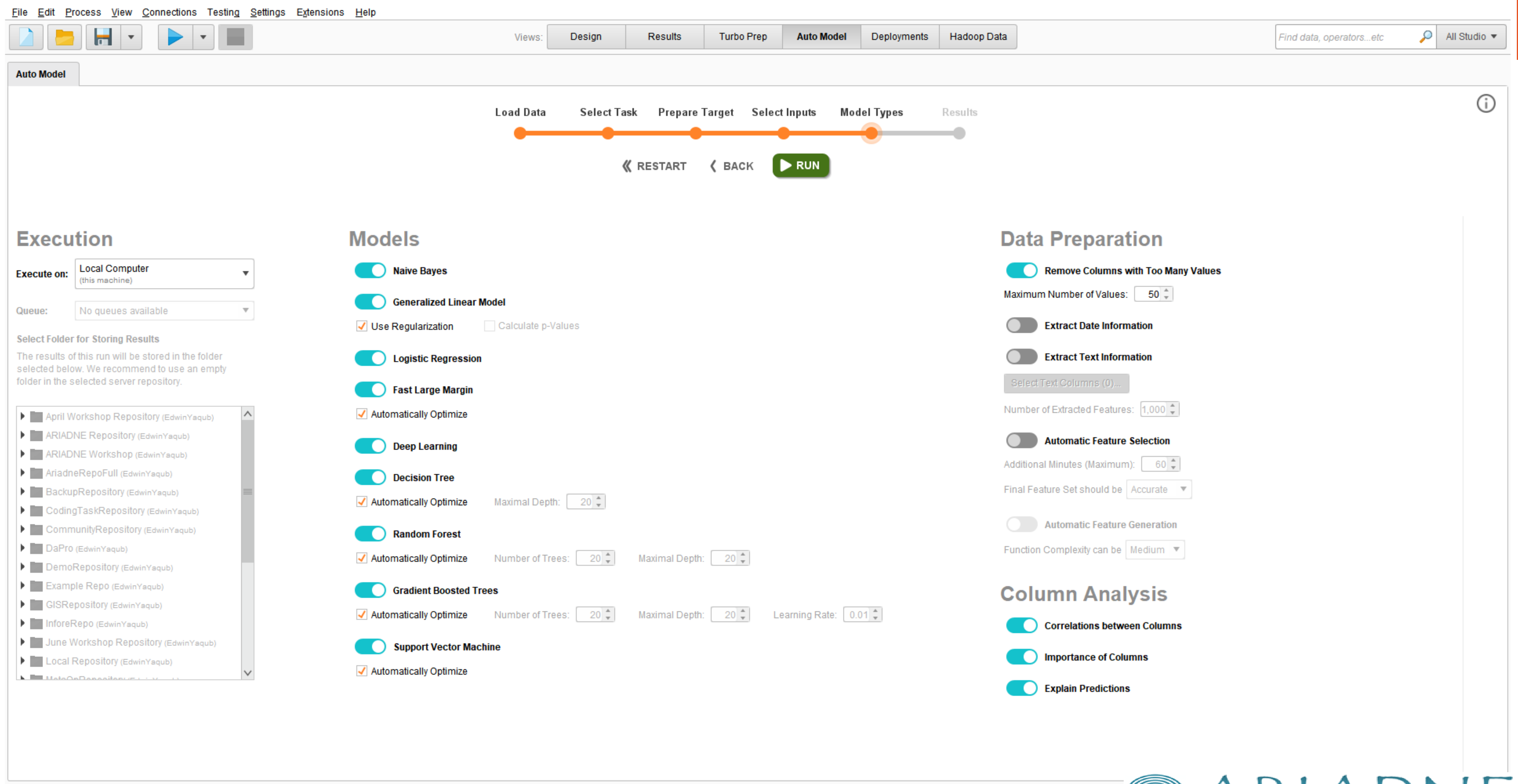
Predict
Want to predict the values of a column?

Clusters
Want to identify groups in your data?

Outliers
Want to detect outliers in your data?

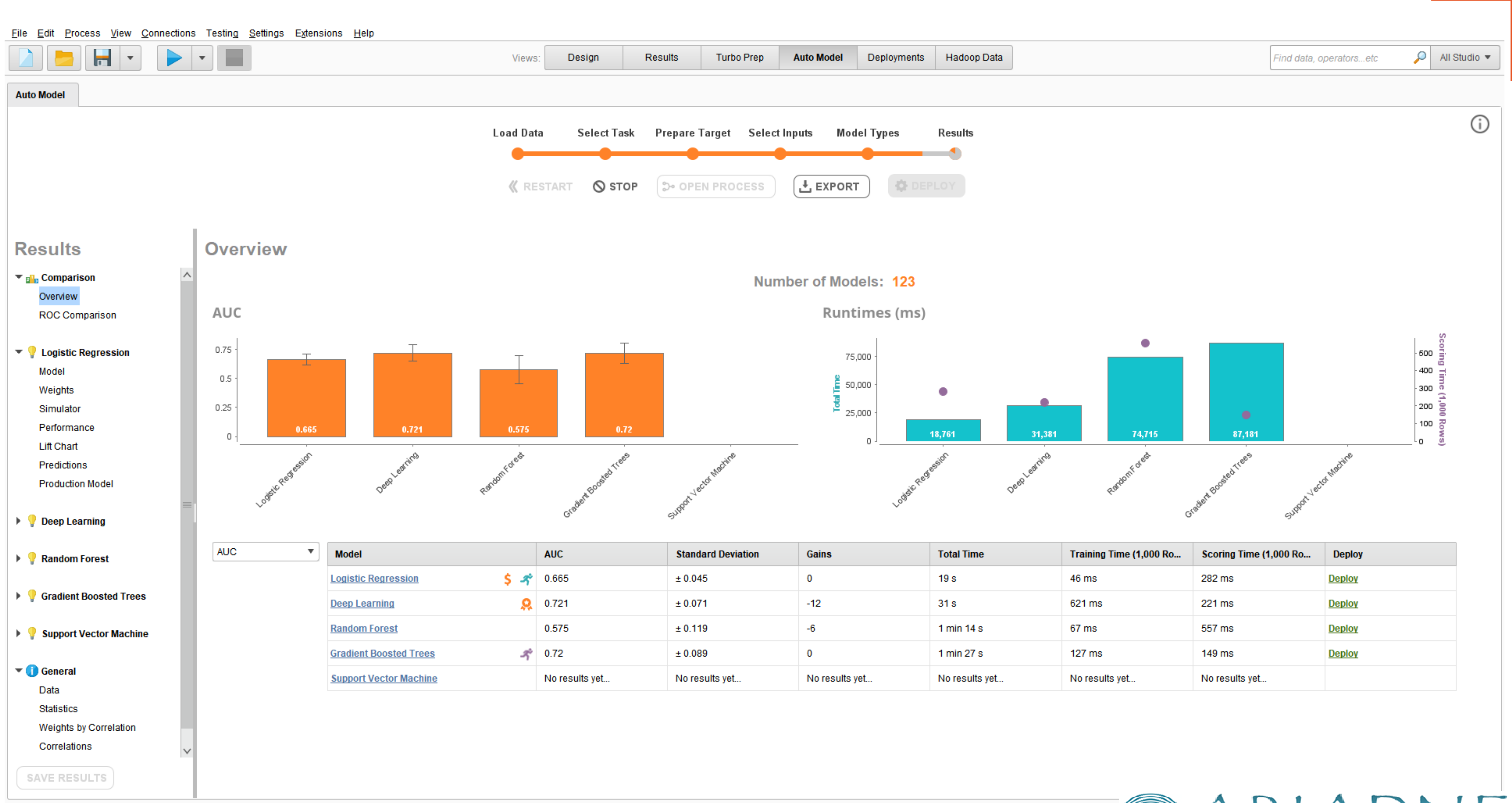
Message	Day Mins (Number)	Day Calls (Number)	Day Charge (Number)	Eve Mins (Number)	Eve Calls (Number)	Eve Charge (Number)	Night Mins (Number)	Night Calls (Number)	Night Charge (Number)	Intl Mins (Number)	Intl Calls (Number)	Intl Charge (Number)	Customer Se... (Number)	State (Category)	Churn Flag (Category)	Customer Ph... (Category)
	184.50000	97	31.37000	351.60000	80	29.89000	215.80000	90	9.71000	8.70000	4	2.35000	1	LA	Loyal	408-335-4719
	141.10000	92	23.99000	249.10000	126	21.17000	136	73	6.12000	10.80000	2	2.92000	2	LA	Loyal	415-382-4024
	163.50000	77	27.80000	203.10000	102	17.26000	232	87	10.44000	7.80000	4	2.11000	2	NY	Loyal	415-391-1348
	220.50000	94	37.49000	239.50000	126	20.36000	254.30000	109	11.44000	5.90000	9	1.59000	2	FL	Loyal	415-357-4936
	118.50000	86	20.15000	213.90000	118	18.18000	132.60000	99	5.97000	13.40000	3	3.62000	2	FL	Loyal	510-394-8504
	112.80000	133	19.18000	199.40000	116	16.95000	142.70000	105	6.42000	10.10000	5	2.73000	0	RI	Loyal	415-360-1776
	88.50000	87	15.05000	178.80000	108	15.20000	228.70000	96	10.29000	11.50000	3	3.11000	2	NY	Loyal	415-398-8588
	215.60000	78	36.65000	195.30000	119	16.60000	194.40000	65	8.75000	3.60000	5	0.97000	1	CO	Loyal	415-400-5984
	221.30000	106	37.62000	267.60000	98	22.75000	111.50000	80	5.02000	9.30000	7	2.51000	0	VT	Loyal	415-401-4052
	160.80000	91	27.34000	155.80000	82	13.24000	254.30000	103	11.44000	8.50000	3	2.30000	1	NE	Loyal	408-347-2378
	180.60000	92	30.70000	190.90000	114	16.23000	295.60000	125	13.30000	10.30000	4	2.78000	1	NE	Churn	415-410-6791
	94.10000	93	16	147.60000	80	12.55000	213.50000	85	9.61000	10.10000	2	2.73000	0	MS	Loyal	415-340-2239
	92.70000	107	15.76000	127.80000	86	10.86000	225.60000	86	10.15000	9.90000	4	2.67000	3	WV	Loyal	415-396-1106
	272.50000	119	46.33000	226.10000	94	19.22000	159.10000	94	7.16000	16.40000	5	4.43000	3	TN	Loyal	415-339-6477
	141.40000	80	24.04000	123.90000	76	10.53000	323.50000	88	14.56000	8.10000	3	2.19000	2	NY	Loyal	510-394-3312
	155.20000	139	26.38000	268.30000	79	22.81000	186.40000	71	8.39000	9.70000	4	2.62000	3	OK	Loyal	510-406-5532
	2.60000	113	0.44000	254	102	21.59000	242.70000	156	10.92000	9.20000	5	2.48000	3	OK	Loyal	510-403-1128

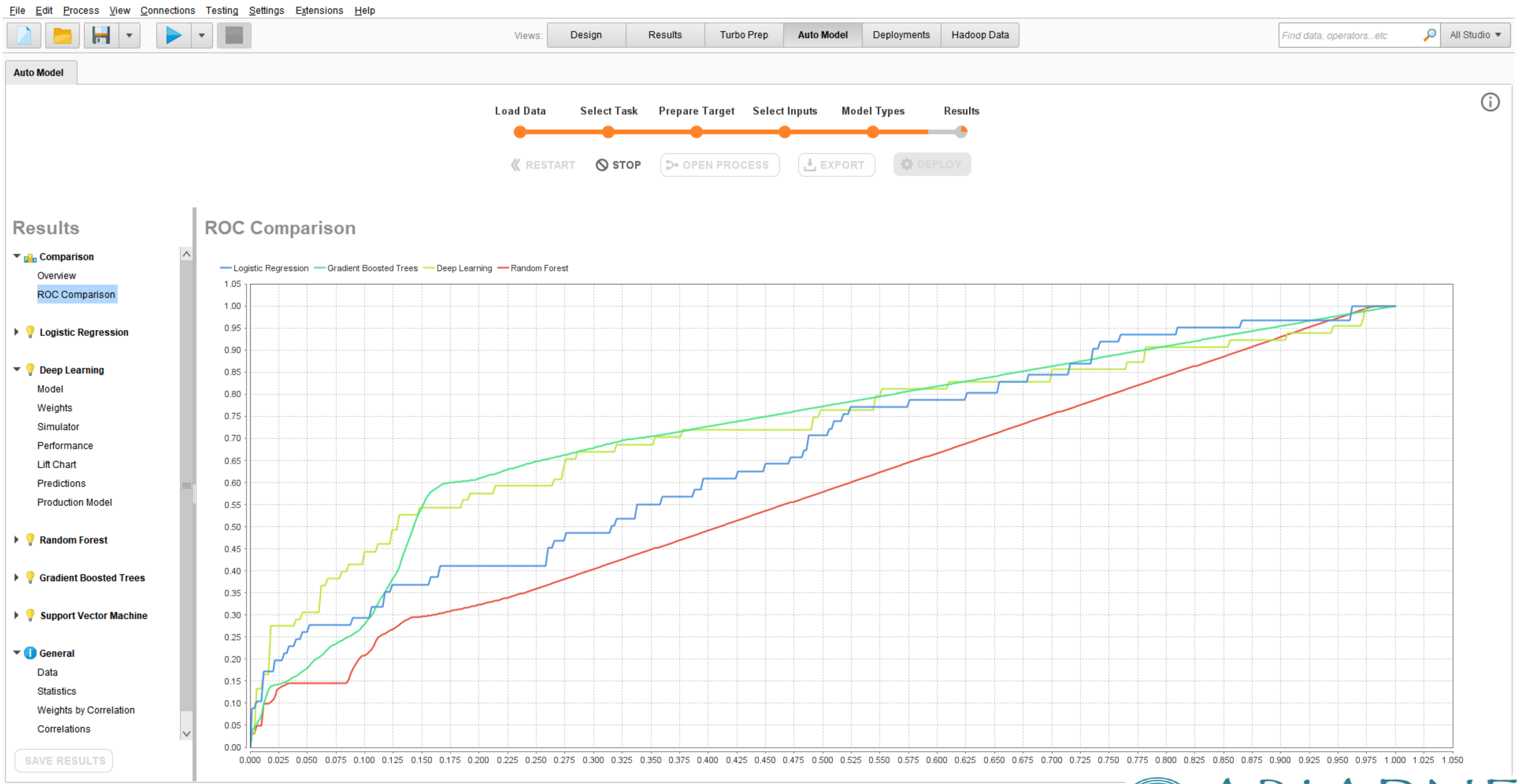
3,333 rows - 20 columns (3 nominal, 15 numerical)

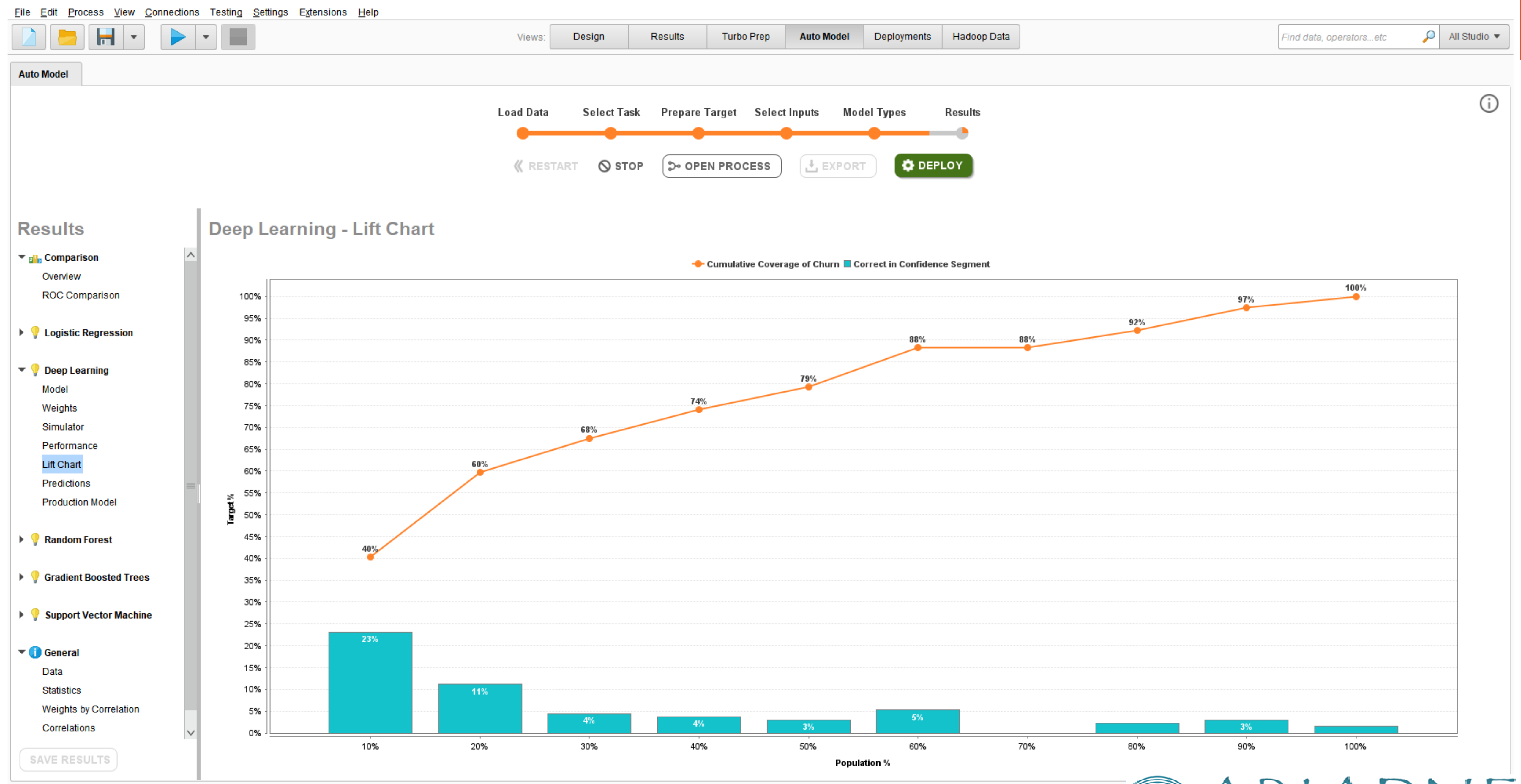


5G PPP

ARIADNE


File Edit Process View Connections Testing Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments Hadoop Data
Find data, operators...etc
All Studio
Auto Model
Load Data Select Task Prepare Target Select Inputs Model Types Results
RESTART STOP OPEN PROCESS EXPORT DEPLOY
Results
Comparison
Overview
ROC Comparison
Logistic Regression
Model
Weights
Simulator
Performance
Lift Chart
Predictions
Production Model
Deep Learning
Random Forest
Gradient Boosted Trees
Support Vector Machine
General
Data
Statistics
Weights by Correlation
Correlations
SAVE RESULTS
Overview
Number of Models: 123
AUC
0.665 0.721 0.575 0.72
Logistic Regression Deep Learning Random Forest Gradient Boosted Trees Support Vector Machine
Runtimes (ms)
Total Time
Scoring Time (1,000 Rows)
18,761 31,381 74,715 87,181
Model | AUC | Standard Deviation | Gains | Total Time | Training Time (1,000 Ro... | Scoring Time (1,000 Ro... | Deploy
Logistic Regression | 0.665 | ± 0.045 | 0 | 19 s | 46 ms | 282 ms | Deploy
Deep Learning | 0.721 | ± 0.071 | -12 | 31 s | 621 ms | 221 ms | Deploy
Random Forest | 0.575 | ± 0.119 | -6 | 1 min 14 s | 67 ms | 557 ms | Deploy
Gradient Boosted Trees | 0.72 | ± 0.089 | 0 | 1 min 27 s | 127 ms | 149 ms | Deploy
Support Vector Machine | No results yet... | No results yet... | No results yet... | No results yet... | No results yet... | No results yet...


File Edit Process View Connections Testing Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments Hadoop Data
Find data, operators...etc
All Studio
Auto Model
Load Data
Select Task
Prepare Target
Select Inputs
Model Types
Results
RESTART
STOP
OPEN PROCESS
EXPORT
DEPLOY
Results
Comparison
Overview
ROC Comparison
Logistic Regression
Deep Learning
Model
Weights
Simulator
Performance
Lift Chart
Predictions
Production Model
Random Forest
Gradient Boosted Trees
Support Vector Machine
General
Data
Statistics
Weights by Correlation
Correlations
SAVE RESULTS
ROC Comparison
Logistic Regression
Gradient Boosted Trees
Deep Learning
Random Forest



5G PPP

ARIADNE

File Edit Process View Connections Testing Settings Extensions Help

Views: Design | Results | Turbo Prep | **Auto Model** | Deployments | Hadoop Data

Find data, operators...etc | All Studio

Auto Model

Load Data · Select Task · Prepare Target · Select Inputs · Model Types · Results

RESTART | STOP | OPEN PROCESS | EXPORT | DEPLOY

Results

- **Comparison**
  - Overview
  - ROC Comparison
- **Logistic Regression**
- **Deep Learning**
  - Model
  - Weights
  - Simulator
  - Performance
  - Lift Chart
  - Predictions
  - Production Model
- **Random Forest**
- **Gradient Boosted Trees**
- **Support Vector Machine**
- **General**
  - Data
  - Statistics
  - Weights by Correlation
  - Correlations

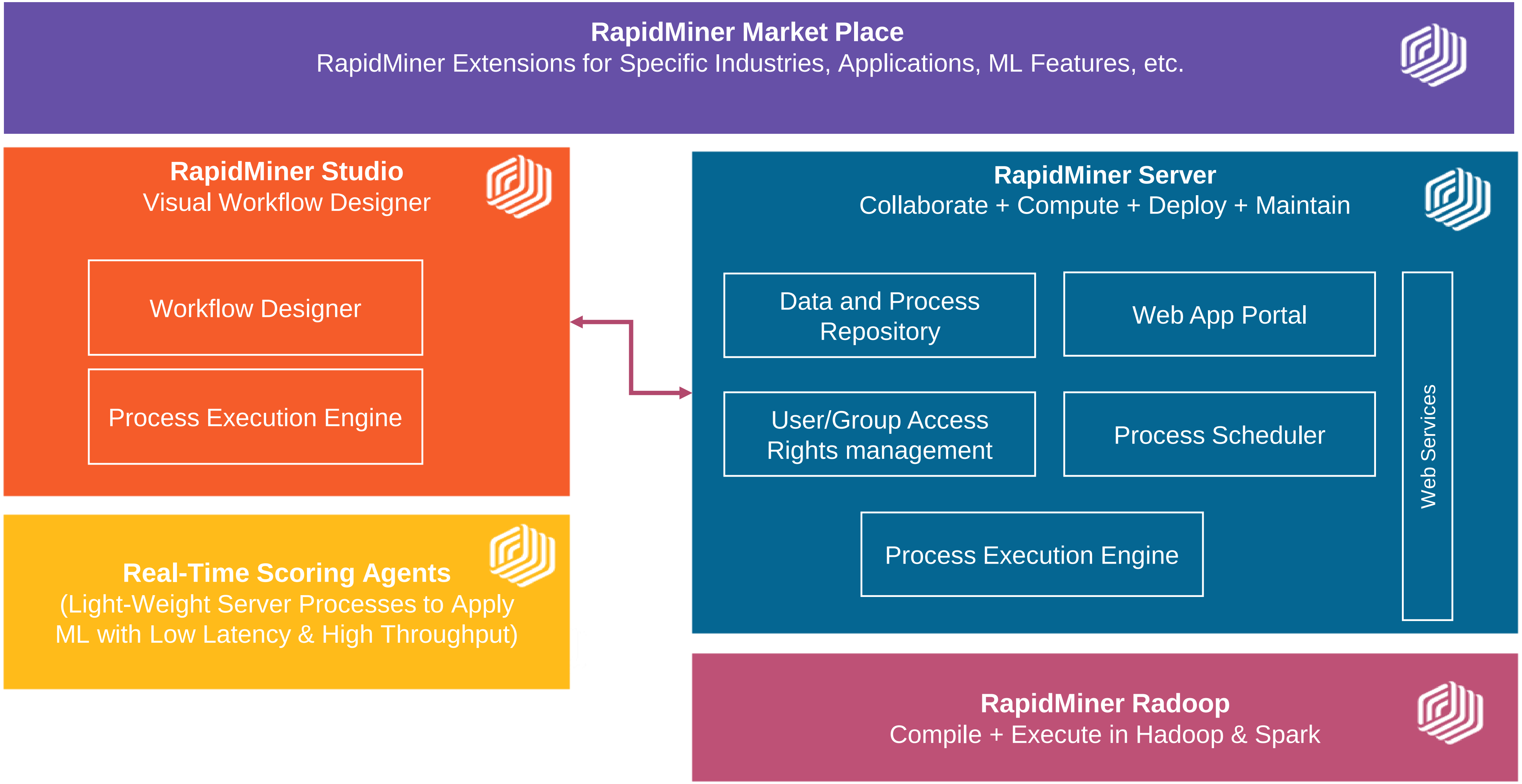
SAVE RESULTS

Deep Learning - Predictions

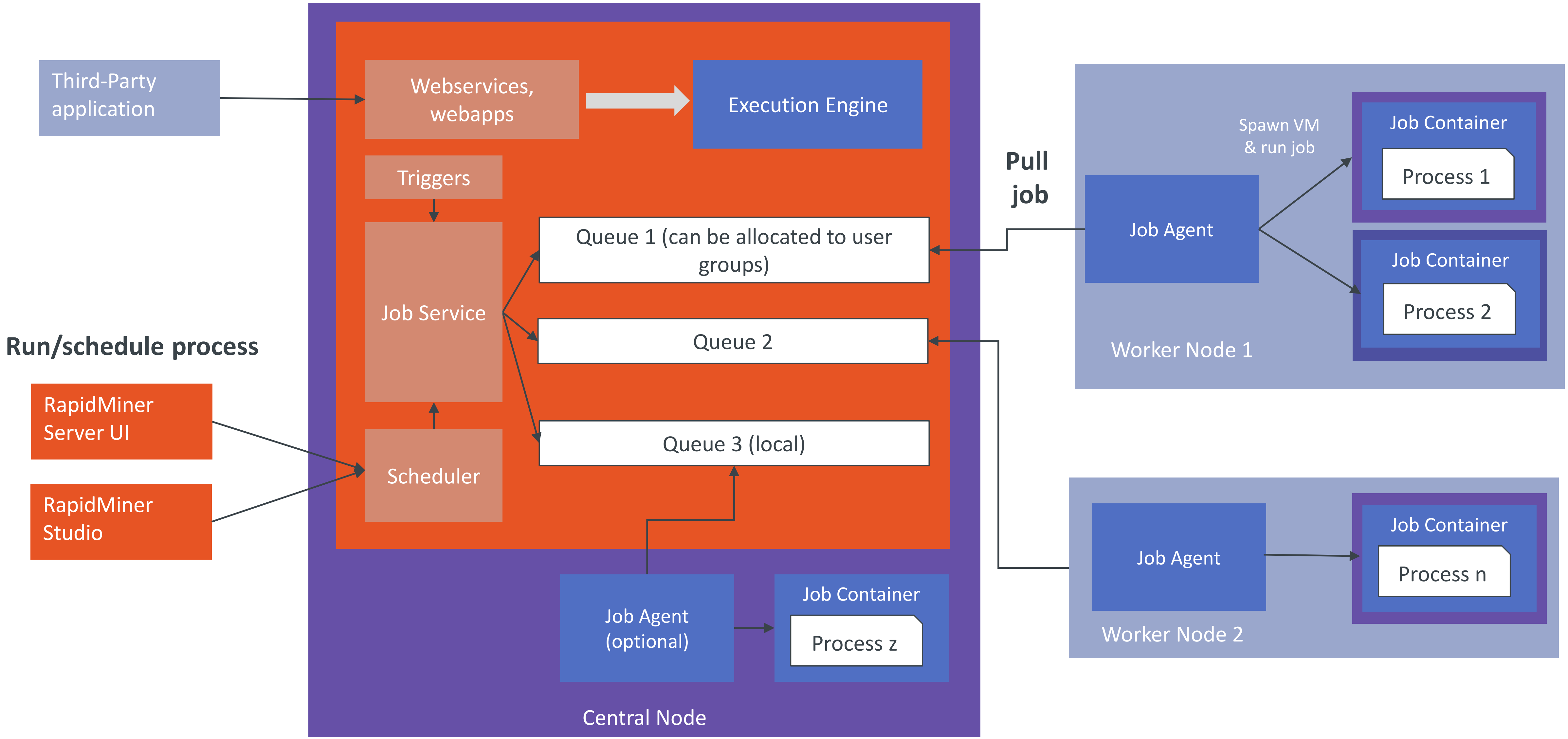
Row No.	Churn Flag	prediction(C...	confidence(...	confidence(...	cost	VMail Plan	Account Len...	VMail Mess...	Day Mins	Day Calls	Day Charge	Eve Mins	Eve Calls	Eve Charge	Night Mins
1	Loyal	Loyal	0.97103	0.02897	0.94206	yes	142	26	220.50000	94	37.49000	239.50000	126	20.36000	254.30000
2	Loyal	Loyal	0.97210	0.02790	0.94421	no	131	0	112.80000	133	19.18000	199.40000	116	16.95000	142.70000
3	Loyal	Loyal	0.96730	0.03270	0.93461	no	108	0	215.60000	78	36.65000	195.30000	119	16.60000	194.40000
4	Churn	Loyal	0.96995	0.03005	0.93990	no	96	0	180.60000	92	30.70000	190.90000	114	16.23000	295.60000
5	Loyal	Loyal	0.74673	0.25327	0.49346	yes	94	28	92.70000	107	15.76000	127.80000	86	10.86000	225.60000
6	Loyal	Loyal	0.59602	0.40398	0.19203	no	112	0	272.50000	119	46.33000	226.10000	94	19.22000	159.10000
7	Loyal	Loyal	0.95818	0.04182	0.91635	no	7	0	206.70000	87	35.14000	281.10000	83	23.89000	158.50000
8	Loyal	Loyal	0.97220	0.02780	0.94439	no	89	0	105.90000	151	18	189.60000	142	16.12000	170.90000
9	Churn	Loyal	0.96799	0.03201	0.93597	no	108	0	115.10000	114	19.57000	211.30000	70	17.96000	136.10000
10	Loyal	Loyal	0.87854	0.12146	0.75708	yes	130	26	257.20000	108	43.72000	224.30000	122	19.07000	204
11	Loyal	Loyal	0.89450	0.10550	0.78900	no	149	0	217.70000	91	37.01000	273.50000	74	23.25000	226.90000
12	Loyal	Loyal	0.96390	0.03610	0.92780	no	146	0	138.40000	104	23.53000	158.90000	122	13.51000	47.40000
13	Churn	Loyal	0.95978	0.04022	0.91956	no	130	0	212.80000	102	36.18000	189.80000	137	16.13000	170.10000
14	Loyal	Loyal	0.97228	0.02772	0.94455	yes	52	24	170.90000	71	29.05000	201.40000	80	17.12000	159
15	Loyal	Loyal	0.96580	0.03420	0.93160	no	70	0	134.70000	96	22.90000	235.90000	90	20.05000	260.20000
16	Loyal	Loyal	0.95132	0.04868	0.90263	no	158	0	209.90000	112	35.68000	221.30000	82	18.81000	210
17	Loyal	Loyal	0.97121	0.02879	0.94243	no	120	0	131.70000	99	22.39000	163.10000	109	13.86000	201.10000
18	Loyal	Loyal	0.97044	0.02956	0.94088	yes	127	22	166	114	28.22000	174.50000	103	14.83000	244.90000
19	Loyal	Loyal	0.97002	0.02998	0.94005	yes	97	28	202.30000	97	34.39000	69.20000	84	5.88000	257.60000
20	Loyal	Loyal	0.81055	0.18945	0.62111	no	103	0	246.50000	47	41.91000	195.50000	84	16.62000	200.50000

Management and Orchestration Tools

RapidMiner Platform: Distributed AI-ready



RapidMiner Server 9.X Architecture



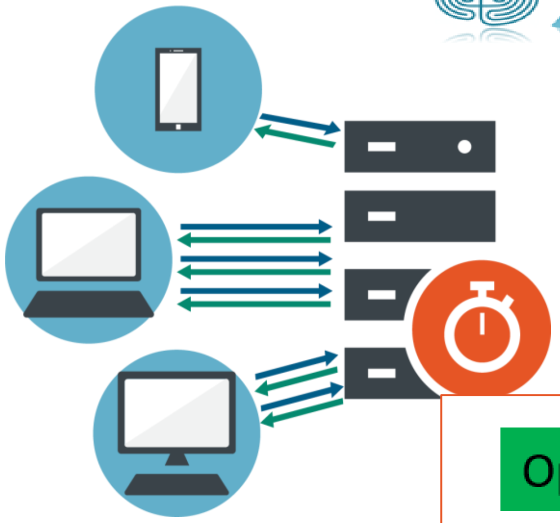
Real-Time Scoring Agents

1) Low memory footprint server processes designed for high throughput low latency (fast scoring) scenarios.

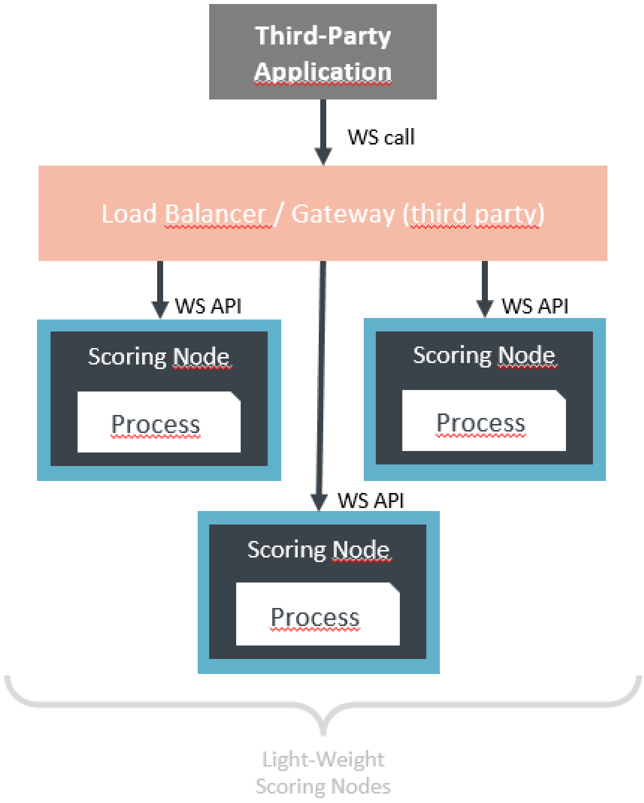
2) Enable real-time online scoring from web portals, phone apps, or desktop applications.

3) Manage deployment of scoring processes as webservices with JSON input/output format

4) Horizontally scalable as docker containers



Operationalization - Scoring



Learning Resources about RapidMiner

RapidMiner Academy: https://academy.rapidminer.com

Online Documentation: https://docs.rapidminer.com

Online Community: https://community.rapidminer.com

More Applications and Use Cases

Marketing

Lead Identification
Recommender
Package Optimization
Cross- & Upsell
Click Route Analysis
Product Tailoring
Mail
E-Mail
Online Ad
Website Optimization
Segmentation
Targeting
Visit Optimization
Police
Care
Sales
Localization / Selection Optimization
Human Resources

Risk Management

Credit Worthiness
Compliance
College dropout
Cart abandonment
Churn
Risks
Customer Retention
Financial Risks
Direct Churn
Fraud
Spam Detection
Bot Detection
Credit card fraud
Money Laundry

Demand Forecasting

Price Optimization
Supply Chain Optimization
Retail
Staff Optimization
IT Forecast
(Cloud Services)
Traffic
Replenishment Prediction
Predictive Packaging
Energy

Manufacturing

Root Cause
IT
Predictive Maintenance
Recipe Optimization
Yield Optimization
Early Rejects
Process Parameter Forecast
Quality assurance

THANK YOU VERY MUCH

Dr. Edwin Yaqub

Senior Data Scientist

eyaqub@rapidminer.com

@byond2day

Ralf Klinkenberg

Founder & Head of Data Science Research

rklinkenberg@rapidminer.com

@RalfKlinkenberg

#1 Agile Predictive Analytics Platform for Today's Modern Analysts