



FADOMAS

Technical Presentation

Vassilis Papataxiarhis
Co-Founder, CEO
vpap@intellia.gr



We automate Predictive Maintenance

Data Provider: **idea75**

the Company


- Est. in 2016
- Data-driven start-up focusing on AI / ML
- Large Portfolio in industrial collaborations
- Predictive maintenance on
 - Ship machinery
 - Material flow logistics




intellia

ML Specialists
& Data Scientists

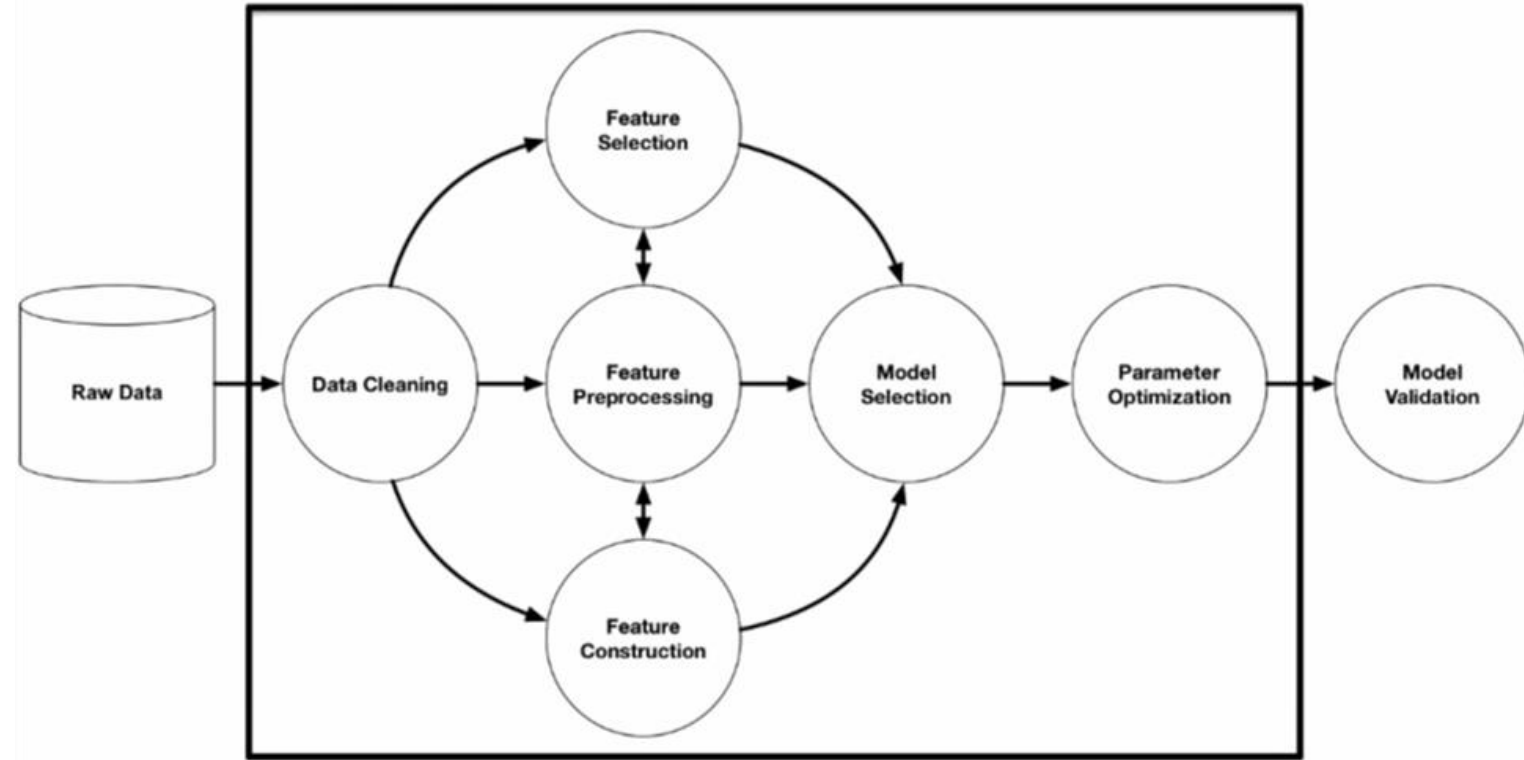
 10+ EU/national
R&D projects

 15+ ML applications

 Research awards
in the interpretation
of complex data

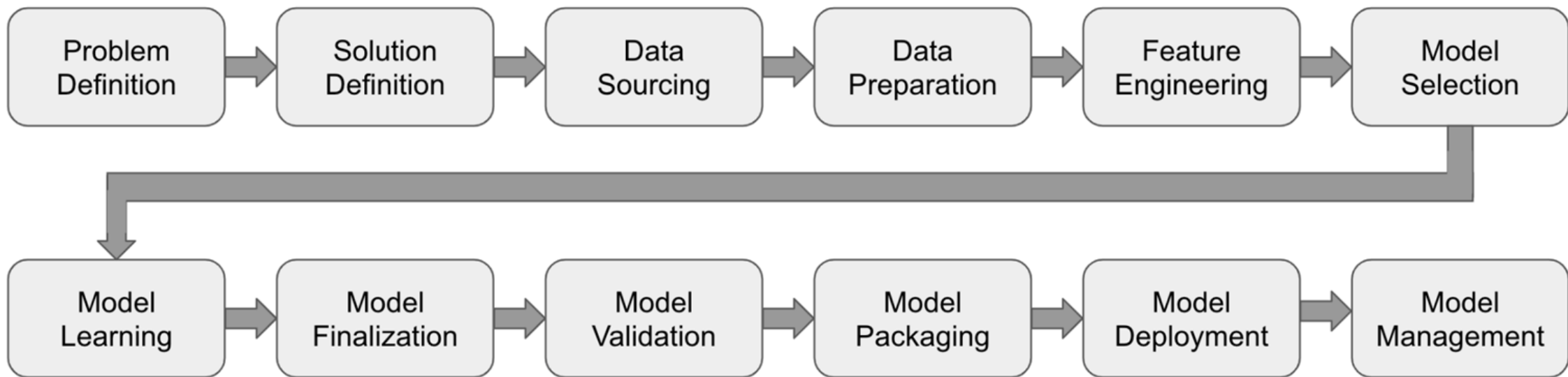


Complexity of ML workflows



source: R. Olson et. al. (2016) "Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science."

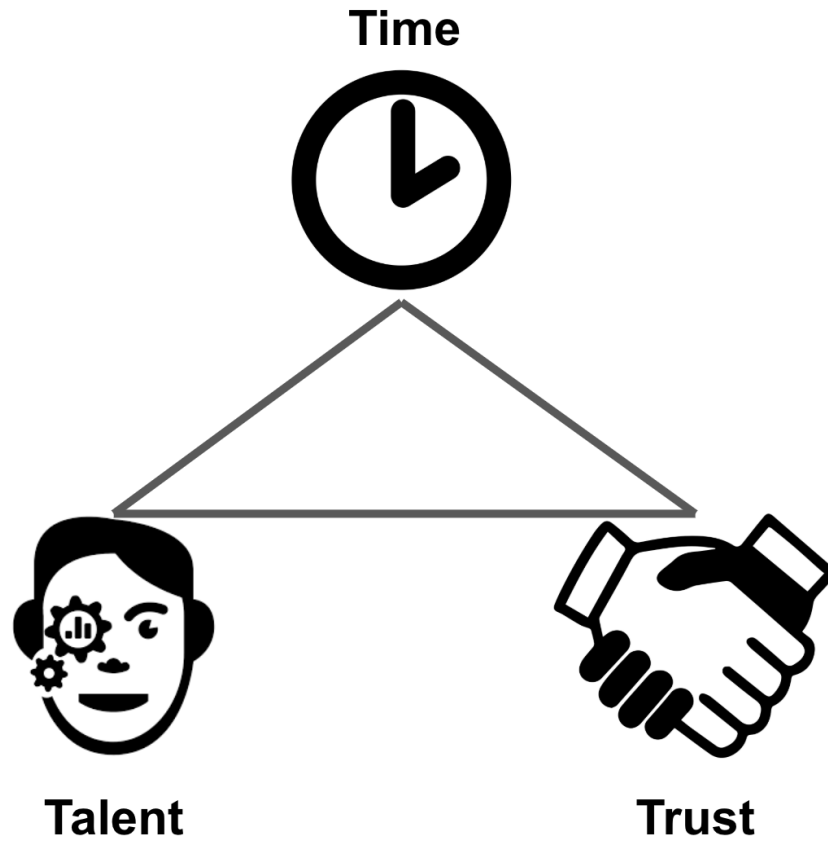




In the **Real-world**

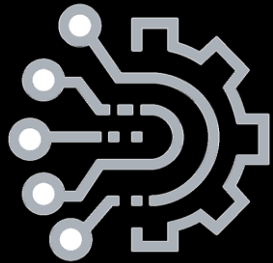


the **Need**



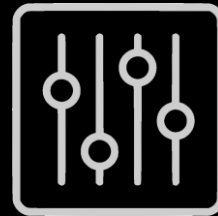
What if we could **simplify ML**?

Automate ...



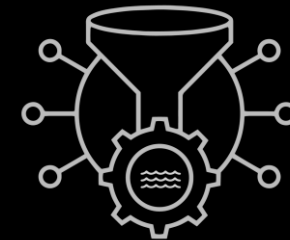
Model selection

Evaluate various models and choose the best one for your data



Configuration

Fine-tune the optimal set of parameters for your model

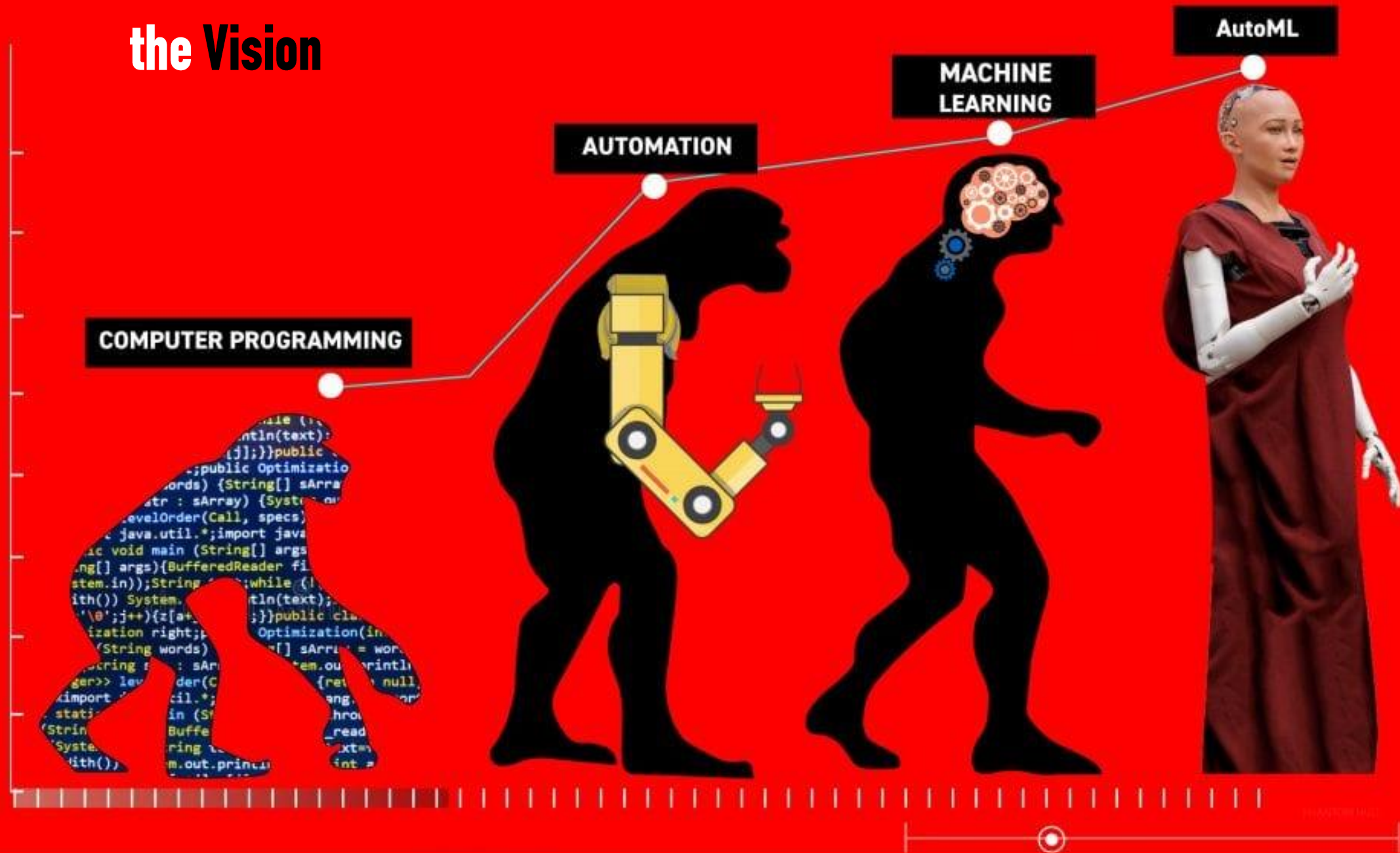


Data Featurization

Data cleaning, Missing Value Substitution, Transform data, etc.

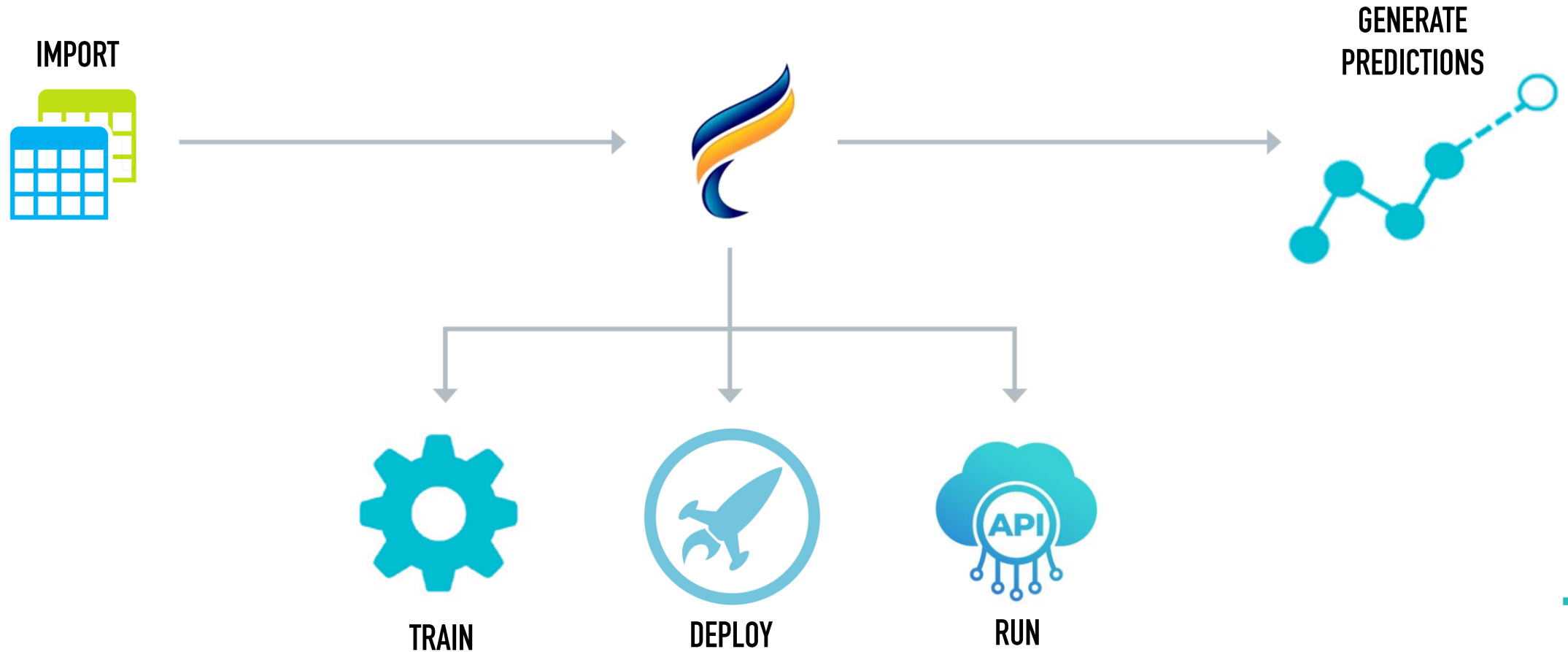


the Vision



FADOMAS SaaS – How it works

fadomise • / fa-ntə-maɪz / verb
hide unnecessary complexity from the users



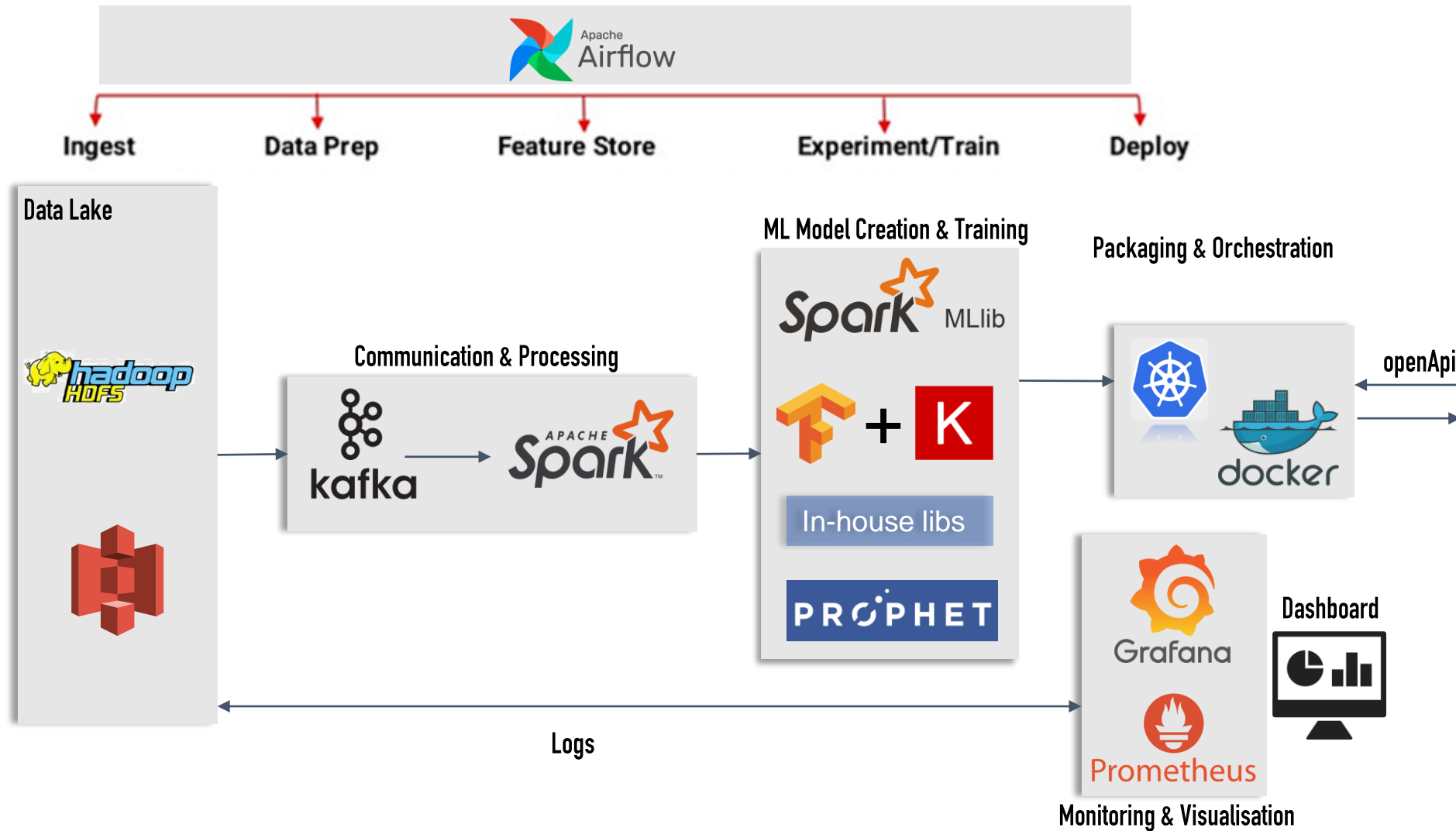
What FADOMAS provides

“FADOMAS starts where rule-based automation ends”

- **Functionality (the ML workflow)**
 - Automated
 - Leaderboard of candidate solutions
- **Versatility (types of problems)**
 - Time-Series forecast, Classification, Anomaly detection
 - Extendable pool of algorithms
- **Flexibility (customization and control power)**
 - Support of industrial standards (IEEE C37.1-2007 for SCADA, OGC SOS / O&M / SensorML)
 - OpenAPI
 - Ability to support different sensor data formats (e.g., JSON, XML)
- **Scalability (large problems in terms of size)**
 - Spark Cluster
 - Kubernetes
 - Microservices Architecture
- **Transparency/Trust (analyze the process)**
 - Model explanation
- **Ease of use (free, GUI, APIs)**



the Architecture



the Algorithms

WILL IT FAIL?	WILL IT FAIL FOR REASON X?	AFTER HOW LONG WILL IT FAIL?	IS THE BEHAVIOUR ANOMALOUS?	WHAT WILL BE THE FUTURE VALUE?
CLASSIFICATION	MULTICLASS CLASSIFICATION	REGRESSION	ANOMALY DETECTION	TIME SERIES FORECASTING
RNN, LSTM	RNN, LSTM	RNN, LSTM	AutoEncoder	Prophet
Deep Neural Classification	Deep Neural Classification	DNN regression	Anomaly MASF	Neural Prophet
Traditional ML: Random Forest, SVM, Decision Trees, GBT	Random Forest Decision Trees, Hidden Markov Chain	Random Forest Regression	Conditional AutoEncoder	
	¹ Partial Matching, Markov Models, Statistical Process Control (CUSUM, Shewhart)			

¹V. Papataxiarhis, S. Hadjiefthymiades, "Event correlation and forecasting over high-dimensional streaming sensor data", 2018 IEEE 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2018), Limassol, Cyprus, October 15-17, 2018.



the **Data Governance**

- **No use of sensitive data BUT industrial data should be secured.**
- **EXPERIMENT phase**
 - **Make use of the REACH infrastructure.**
 - **Data used only for training**
 - **Data visibility and control at all steps.**
 - **Authorised access to data**
 - **Compliance with EU regulations including GDPR**
- **As a next step**
 - **Attribute-based usage control**
 - **Become part of the Federated GAIA-X ecosystem**
 - **Validate DIN SPEC 27070 standardised Reference Architecture with clients**



Quality Assurance & Risk Assessment

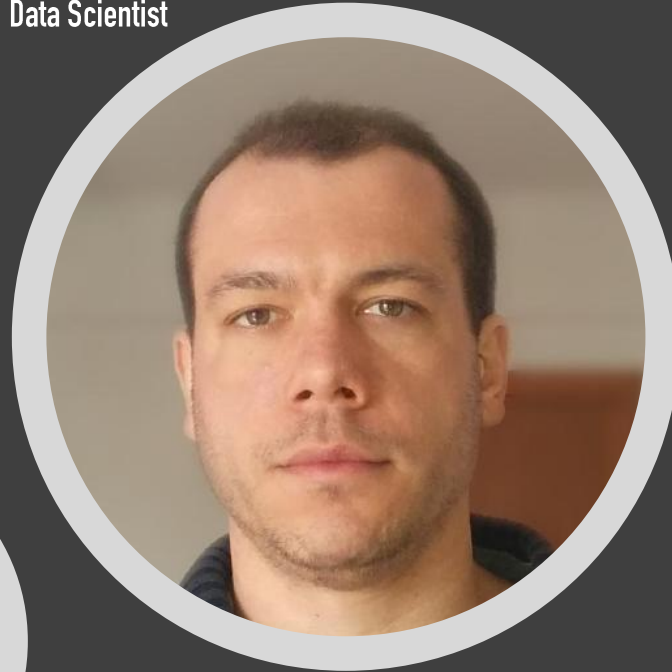
- Agile Development
- Automated CI/CD
 - static analysis
 - exhaustive testing
 - coverage
 - security checks
- Profiling & Logging
- Monitoring
- Detailed documentation

Risk	Low quality of results – overfitting/underfitting				
Impact	High	Likelihood	Low	Phase	TEST
Mitigation	Extend toolkit with appropriate algorithms/tools, extend experimentation and testing to modify default parameters/hyper-parameters.				
Risk	High response times may result in outdated predictions				
Impact	High	Likelihood	Medium	Phase	DESIGN
Mitigation	Extend forecasting, exploit queueing system to improve performance.				
Risk	Inability to handle (very) Big Data – System crash				
Impact	Medium	Likelihood	Medium	Phase	DEPLOY
Mitigation	Scale-up policies/load balancing strategies. Use of processing/storage resources is re-examined. Dimensionality reduction may take place.				





Marilena Athanasiou —
Data Scientist



Dimitris Katris —
Data Engineer, Researcher



Sofia Kostakonti — DevOps



Panos Papadatos —
Infrastructure support



Vassilis Papataxiarhis —
co-Founder, CEO

the **Team**



Michael Loukeris —
co-Founder, Marketing



Service Name*

Description

Next



Add New Service



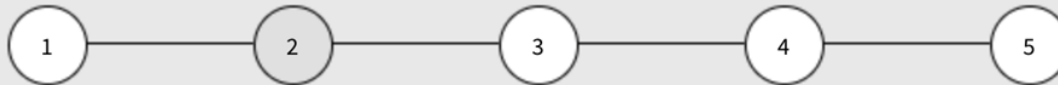


Import your Dataset

- Upload file from your Computer
- Import data from Cloud Storage



Next





Visualisation and Configuration

Summary

Number of Columns: 12
 Number of Rows: 50K
 Total Size: 100MB

Target Column

Current

[Edit additional features](#)

#	Dispositivo	Parametro	Data	Voltage	Active Ene...	Active Power	Current
1	e745f62896fb7909fdd347b4f721...	Valore medio ...	16:30:04 12/10/2020	205.8	471.270	12.89	27.51
2	e745f62896fb7909fdd347b4f721...	Valore medio ...	16:31:04 12/10/2020	206.0	471.270	12.97	27.67
3	e745f62896fb7909fdd347b4f721...	Valore medio ...	16:32:04 12/10/2020	206.2	471.270	13.44	28.39
4	e745f62896fb7909fdd347b4f721...	Valore medio ...	16:33:04 12/10/2020	206.0	471.270	13.21	27.88
5	e745f62896fb7909fdd347b4f721...	Valore medio ...	16:34:04 12/10/2020	205.8	471.270	13.24	28.05

Column Names	Data type	count	mean	std	min	max	Missing Values	histogram
Dispositivo	categorical	30	-	-	-	-	0	-
Parametro	categorical	30	-	-	-	-	0	-
Data	datetime	30	-	-	-	-	0	-
Voltage	numerical	30	205.7	205.8	204.8	208.8	0	show
Active Energy	numerical	30	471.274	471.270	471.270	471.276	0	show
Active Power	numerical	30	13.06	13.04	12.89	13.44	0	show
Current	numerical	30	27.44	27.55	27.51	28.39	0	show

[Next](#)

1

2

3

4

5

Configure





Model Training

Classification

Time Series Forecasting

Edit Parameters

Training Cost

You only pay for hours used: if your model stops improving, training will stop.

The time required to train your model depends on the size and complexity of your training data.

START TRAINING





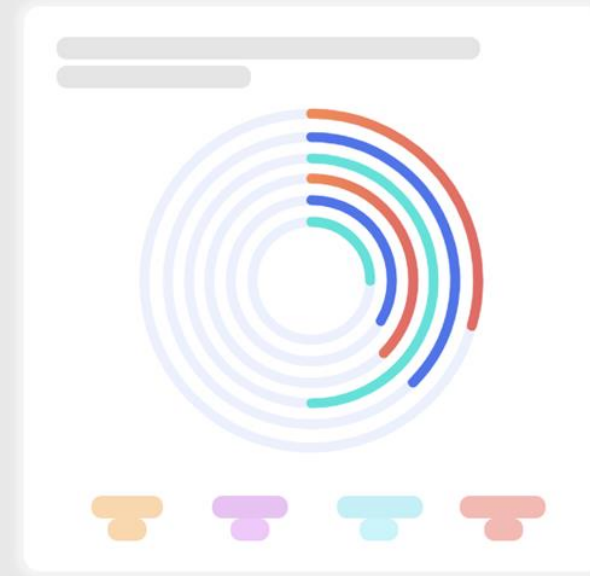
Evaluation

Rank	Model Name	Accuracy	Precision	Recall	AUC ROC	Duration
1	Partial Matching	98.0	95.10	92.10	0.74	5s
2	Gradient Boosted Decision Trees	97.0	95.10	92.10	0.74	5s
3

Based on your preferred optimisation option

Accuracy

[Additional Info](#)



DEPLOY



Evaluation





DEPLOYMENT

Well done! You successfully deploy your ML model! ✕

REST API

request.json

```
{  
  "data": {  
    "Dispositivo": "e745f62896fb7909fdd347b4f721ce24e0222160a3769b5b.....",  
    "Parametro": "Valore medio di corrente",  
    "Data": "16:31:04 12/10/2020",  
    "Valore": 28.300  
  }  
}
```

```
curl -X POST -H "Content-Type: application/json" \  
  -H "Authorisation: Bearer YOUR_ACCESS_TOKEN" \  
  https://mynewcontainer.fadomas.com/projects/models/predict \  
  -d @request.json
```

VISUALIZE





MONITORING: my_service

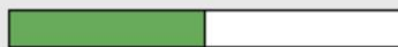
Status

Total Requests
per month

1001

Estimated Cost
per month

100\$

Model
Healthness

RETRAIN

Model Info

*Algorithm: Partial Matching**Optimised for: Accuracy**Value: 98%**First Deployed: 19/5/2021*Mean Response
Time

100 msec

Predictions

2020-10-12 17:01:04

Normal

2020-10-12 17:02:04

Warning Level 1

2020-10-12 17:03:04

Warning Level 1

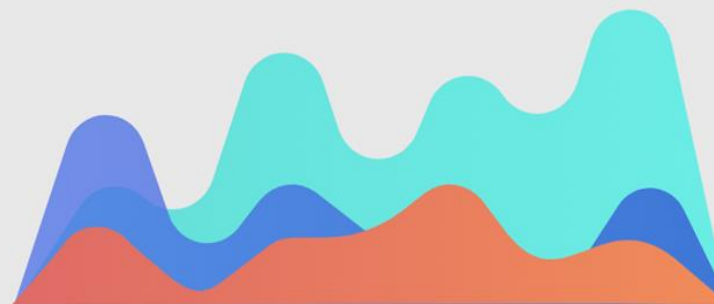
2020-10-12 17:04:04

Warning Level 1

2020-10-12 17:05:04

Warning Level 1

Response Time





MONITORING: my_service

Status



Total Requests per month

1001

Estimated Cost per month

100\$

Model Healthness



RETRAIN

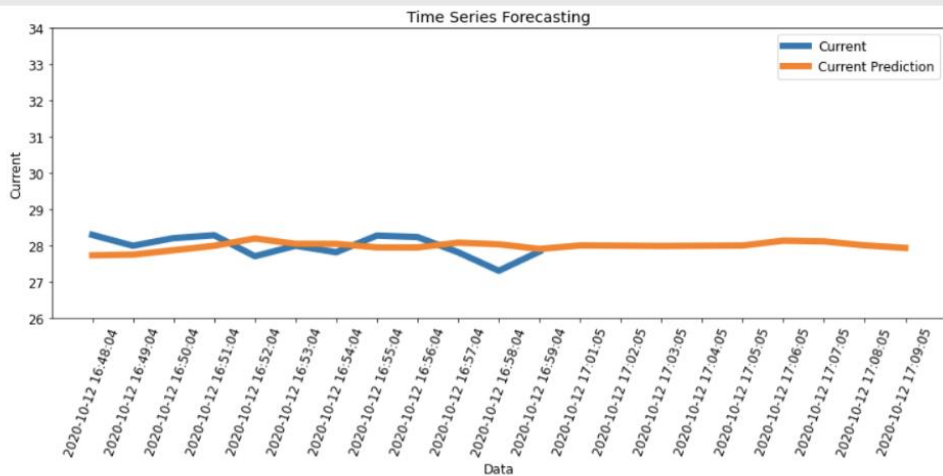
Model Info

Algorithm: Prophet
 Future Prediction: 9 periods
 First Deployed: 19/5/2021

Mean Response Time

100 msec

Predictions



Response Time



Thank you.



Vassilis Papataxiarhis
vpap@intellia.gr