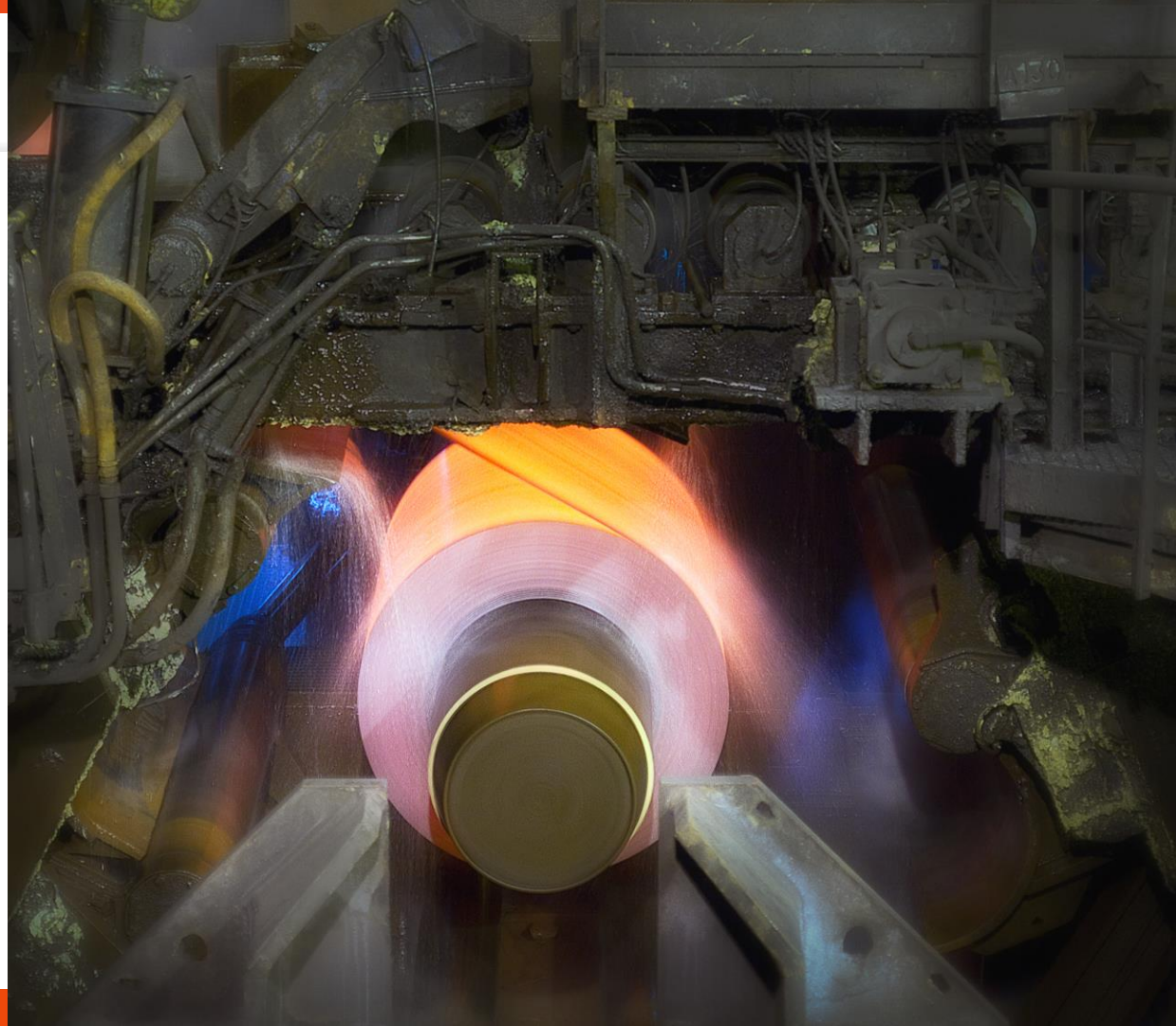
rapidminer

A Reference Architecture for Quality Improvement in Steel Production

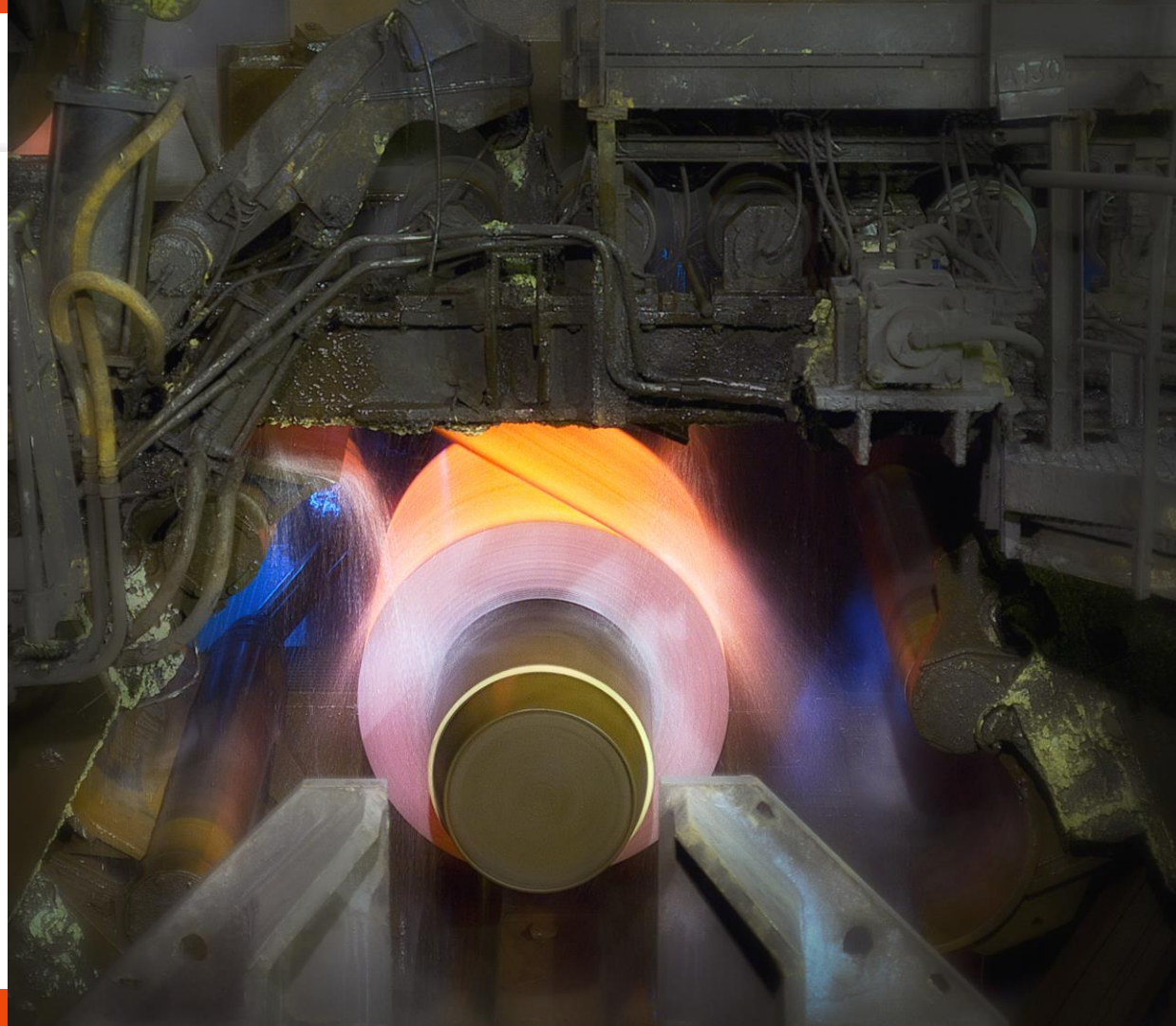
Edwin Yaqub, David Arnu

iDSC 2017 Salzburg, Austria



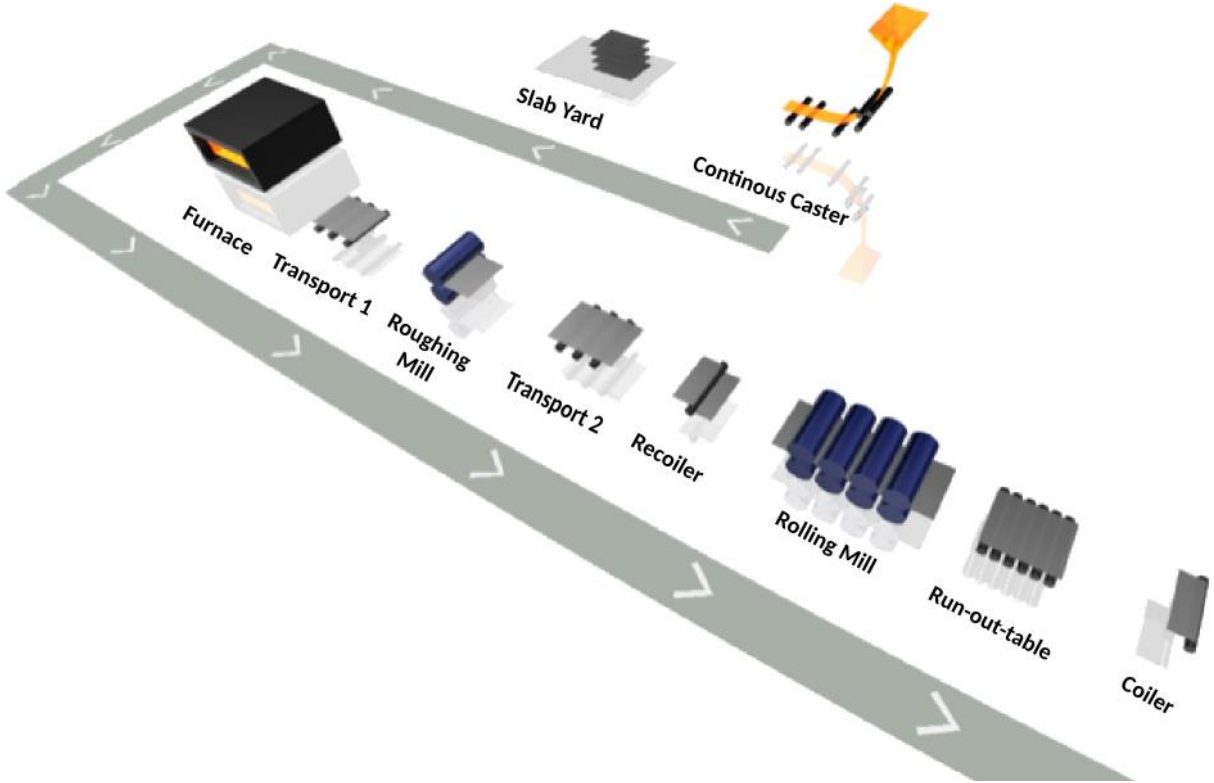
Fos-sur-Mer, France. All rights acquired provided photo credits are mentioned: Alain Sauvan / ArcelorMittal Fos-sur-Mer

- Global demand
- Costs
  - Energy
  - Time
  - Financial
- Data
  - High variability
  - Large



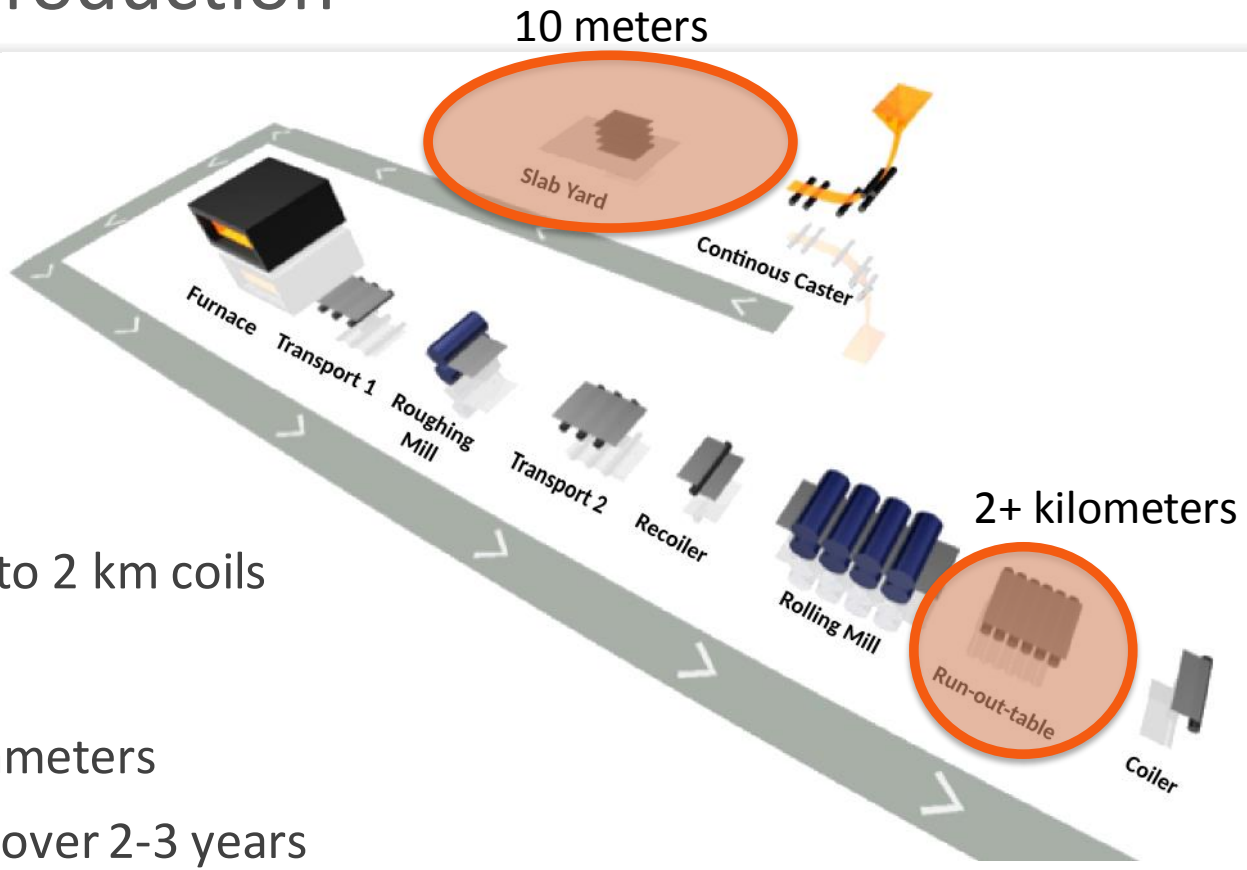
Fos-sur-Mer, France. All rights acquired provided photo credits are mentioned: Alain Sauvan / ArcelorMittal Fos-sur-Mer

Stages in Steel Production



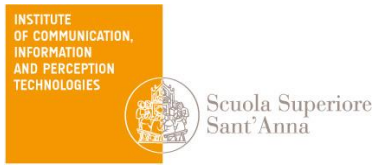
Stages in Steel Production

- Malleable product
- Changing properties
  - From 10 meters slab to 2 km coils
- Lot of raw sensor data
  - Several hundred parameters
  - Frequency of 1-10Hz over 2-3 years



PRESED

Predictive Sensor Data Mining for Product Quality Improvement



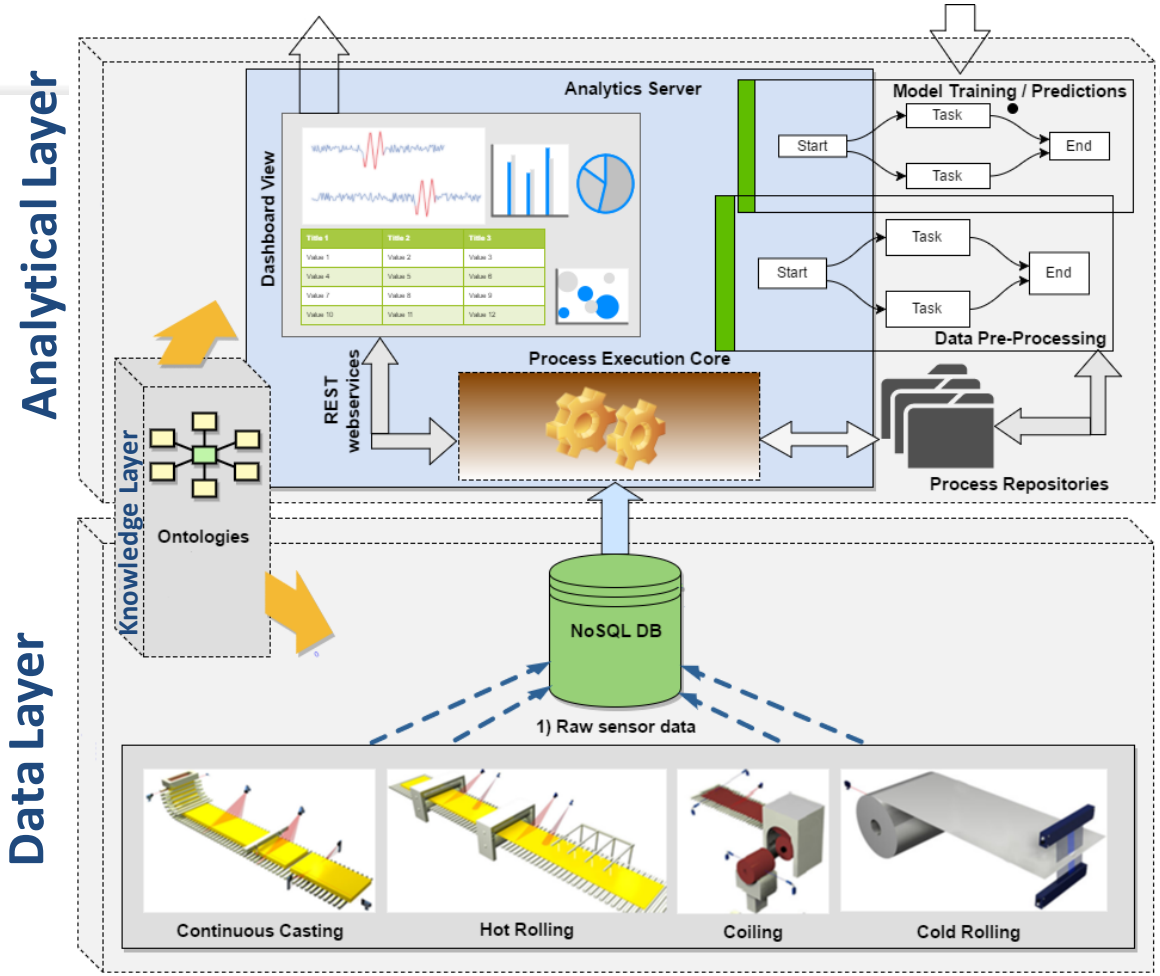
http://www.presed.eu/

PRESED Objectives

- Improve quality prediction / detect defects
- Product orientation concept
- Machine Learning for steel production
  - Information gain (fine granularity)

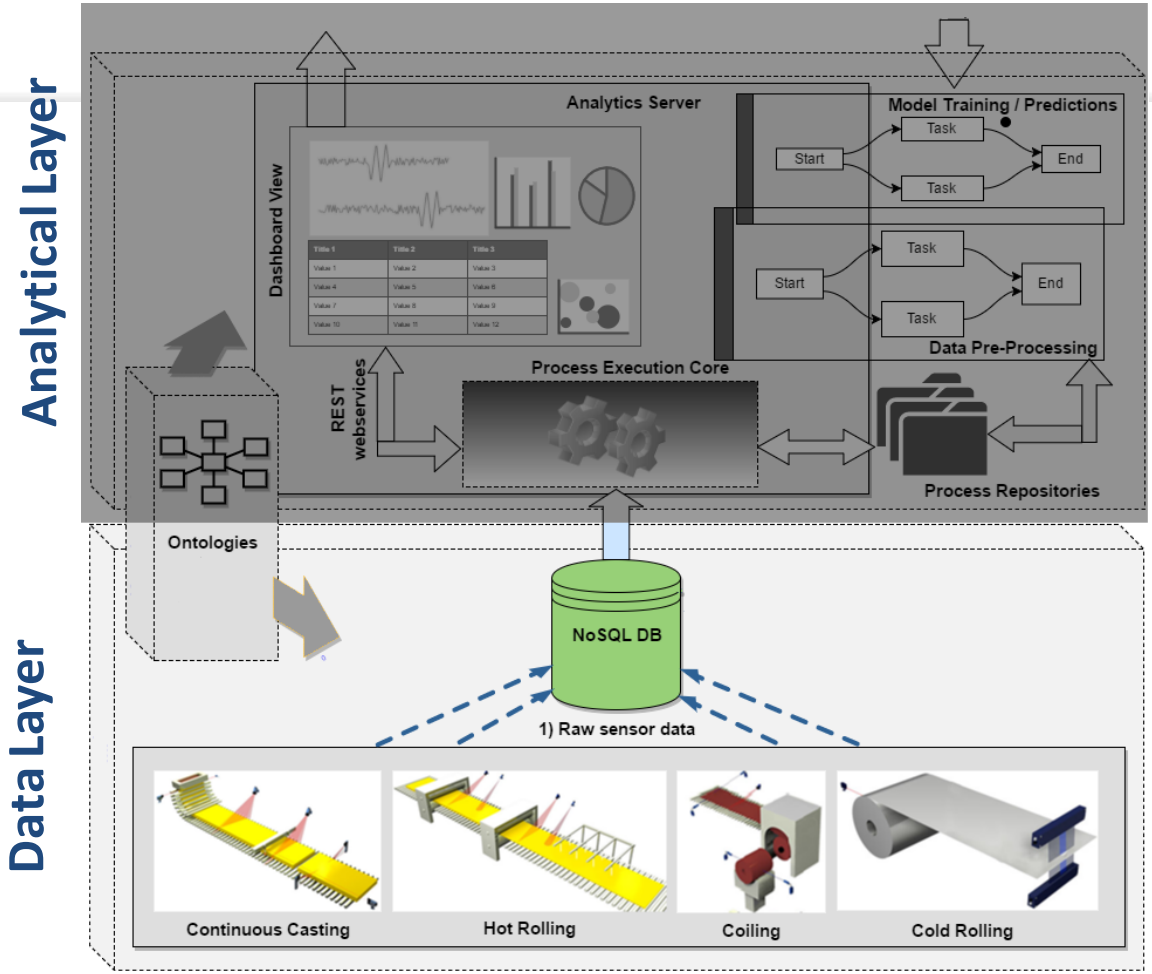
A reference architecture to achieve these objectives

Plant Operators/
Manager



Domain Experts

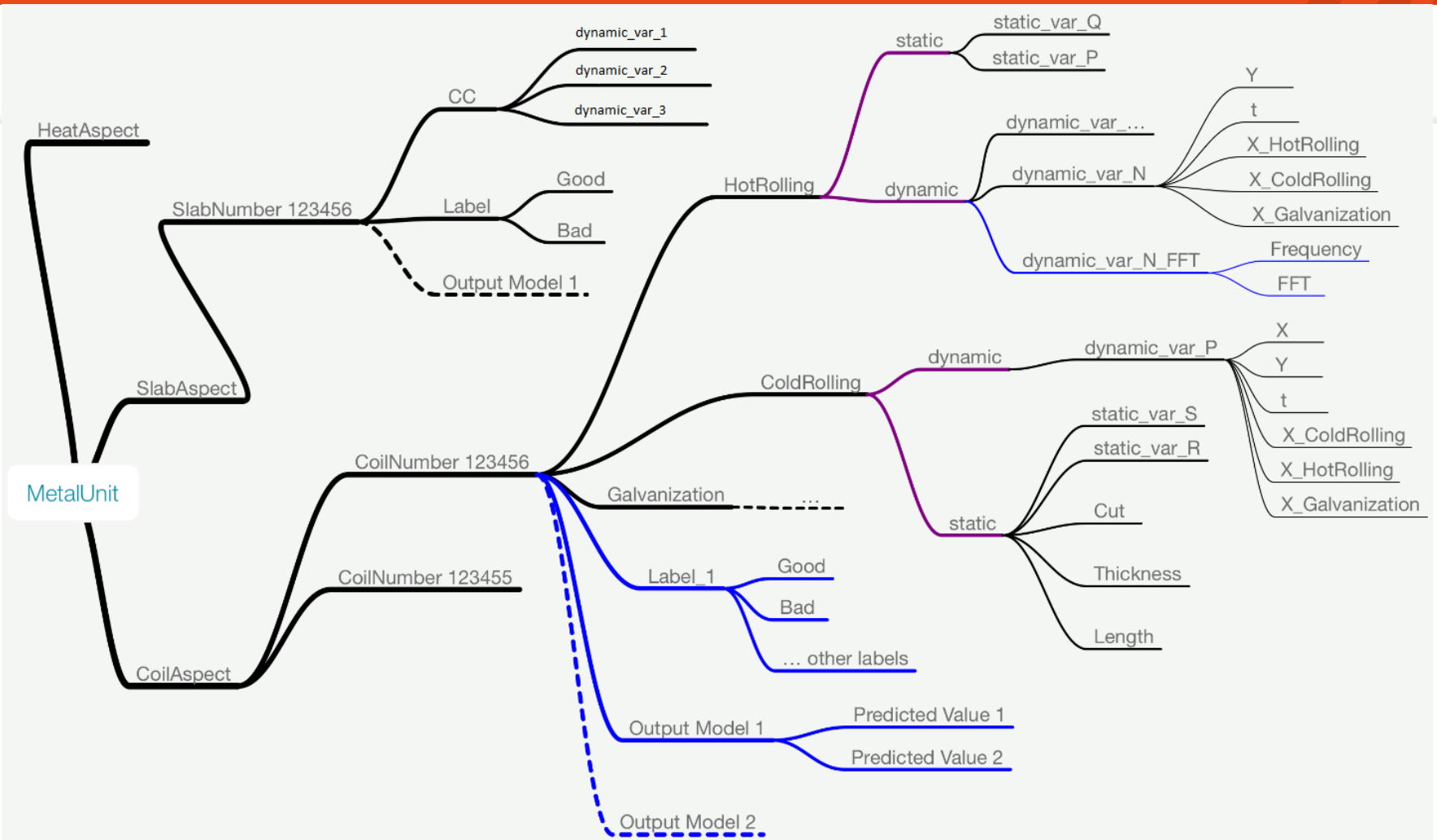
Data Engineers



Data Engineers

Metal Unit Approach (Data Layer)

- Data model for production sensor data
- Object oriented (using NoSQL)
  - Malleable data structure
- Hierarchically organized to fit different production steps

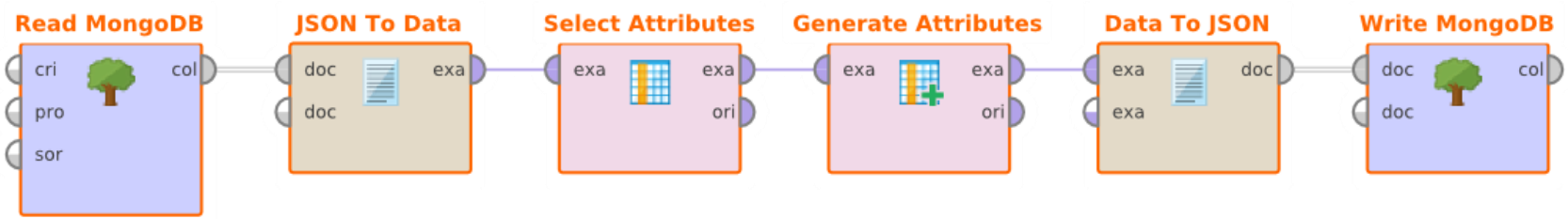

MetalUnit
HeatAspect
SlabAspect
SlabNumber 123456
CC
dynamic_var_1
dynamic_var_2
dynamic_var_3
Label
Good
Bad
Output Model 1
CoilAspect
CoilNumber 123456
CoilNumber 123455
HotRolling
static
static_var_Q
static_var_P
dynamic
dynamic_var_...
dynamic_var_N
Y
t
X_HotRolling
X_ColdRolling
X_Galvanization
dynamic_var_N_FFT
Frequency
FFT
ColdRolling
dynamic
dynamic_var_P
X
Y
t
X_ColdRolling
X_HotRolling
X_Galvanization
static
static_var_S
static_var_R
Cut
Thickness
Length
Galvanization
...
Label_1
Good
Bad
... other labels
Output Model 1
Predicted Value 1
Predicted Value 2
Output Model 2

Tools of Choice (NoSQL)

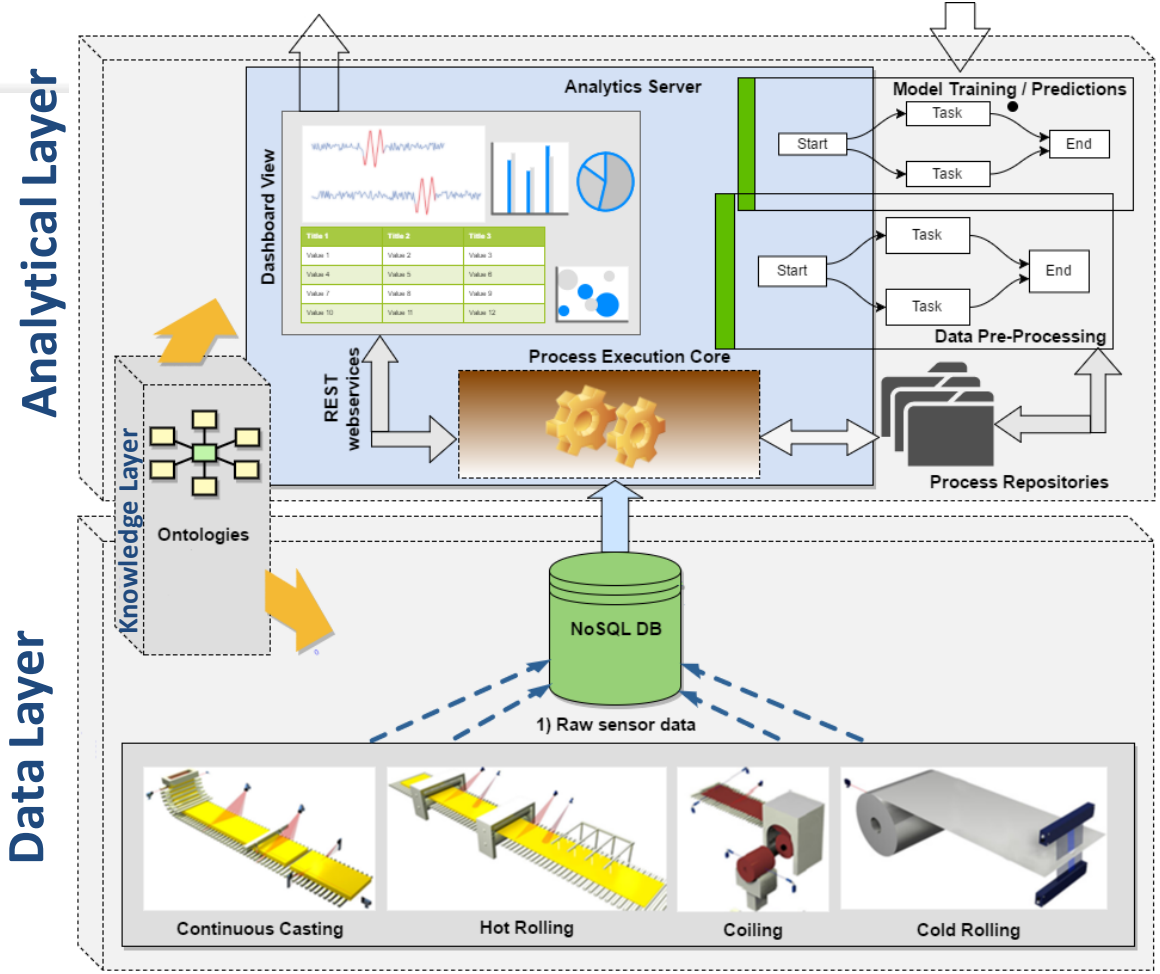
- MongoDB features
- Key-Value Pairs
- JSON format


```
{
    "_id" : ObjectId("58d92891e61483298447fb35"),
    "att1" : 3.30822103287314,
    "att2" : 9.74272379131398,
    "att3" : 1.09643316450764,
    "Label" : true,
    "Orientation" : 0,
    "dynamic_var_1": [
                    0.44934690256,
                    0.44934690256,
                    0.483399136493,
                    0.572202021064,
                    0.661672596495,
                    0.750475481066,
                    0.767048102921
     ]
    ...
    ...
}
```

Tools of choice (RapidMiner)



Plant Operators/
Manager

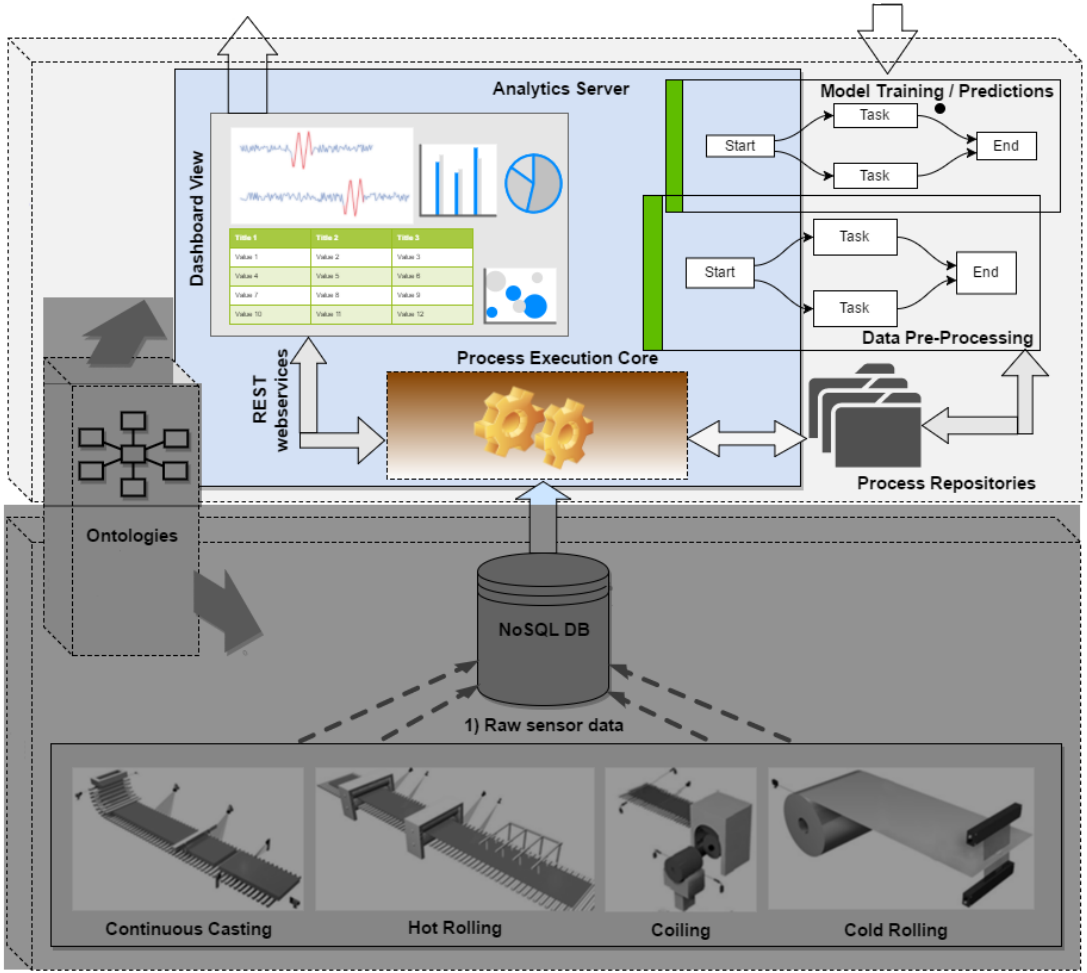


Domain Experts

Data Engineers

Plant Operators/
Manager

Analytical Layer



Domain Experts

Analytics Layer (Smart data)

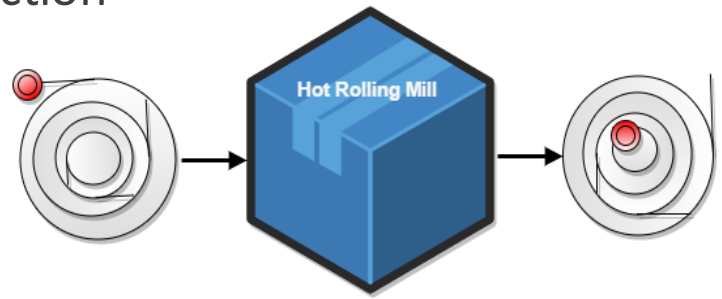
- Smart Data through data enrichment
  - Signal representation (FFT, filtering, differentiation)
  - Feature generation (statistics, Shapelets)

- Calculate rescaling
  - Track data changes during production
  - Changing physical properties

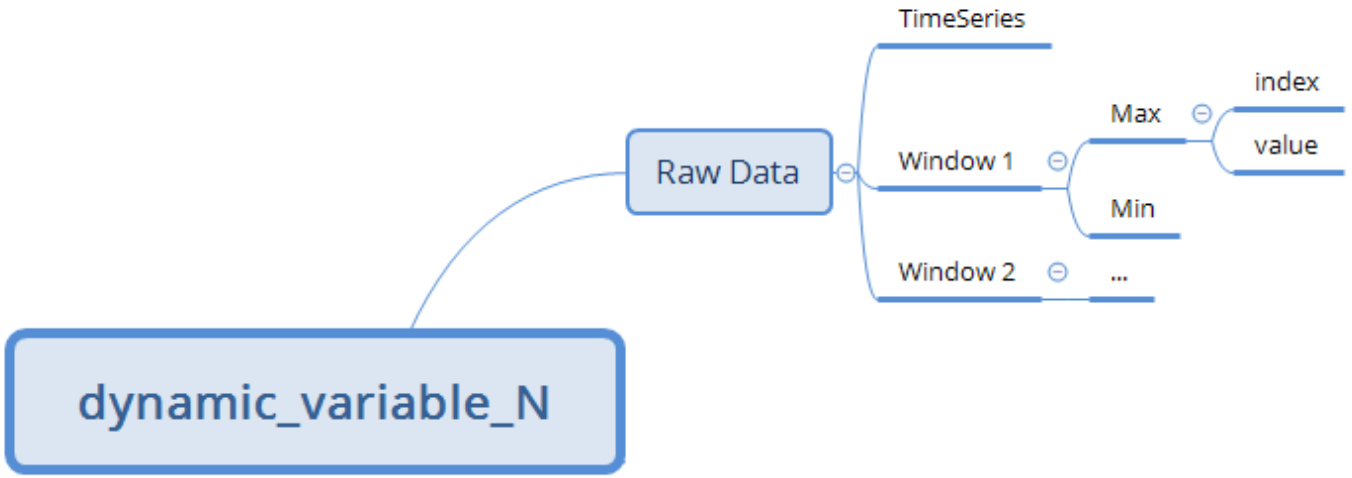
Analytics Layer (Smart data)

- Smart Data through data enrichment
  - Signal representation (FFT, filtering, differentiation)
  - Feature generation (statistics, Shapelets)

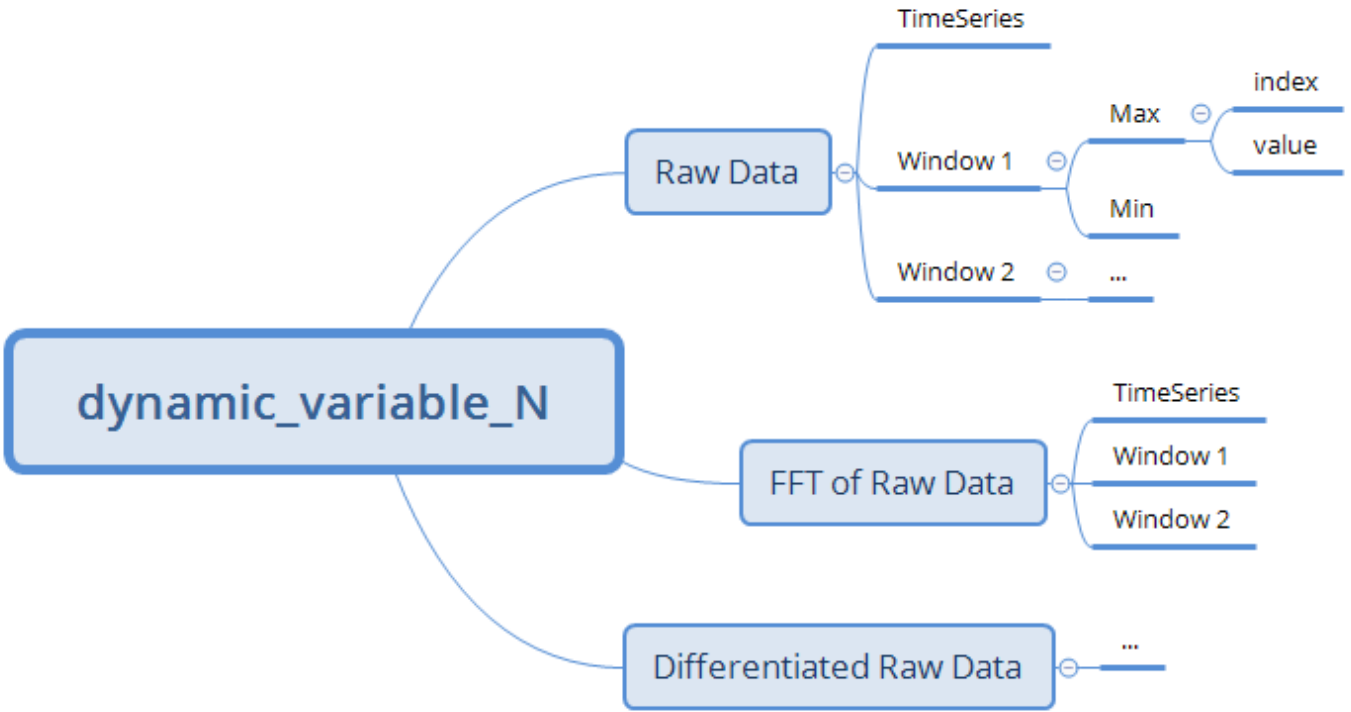
- Calculate rescaling
  - Track data changes during production
  - Changing physical properties



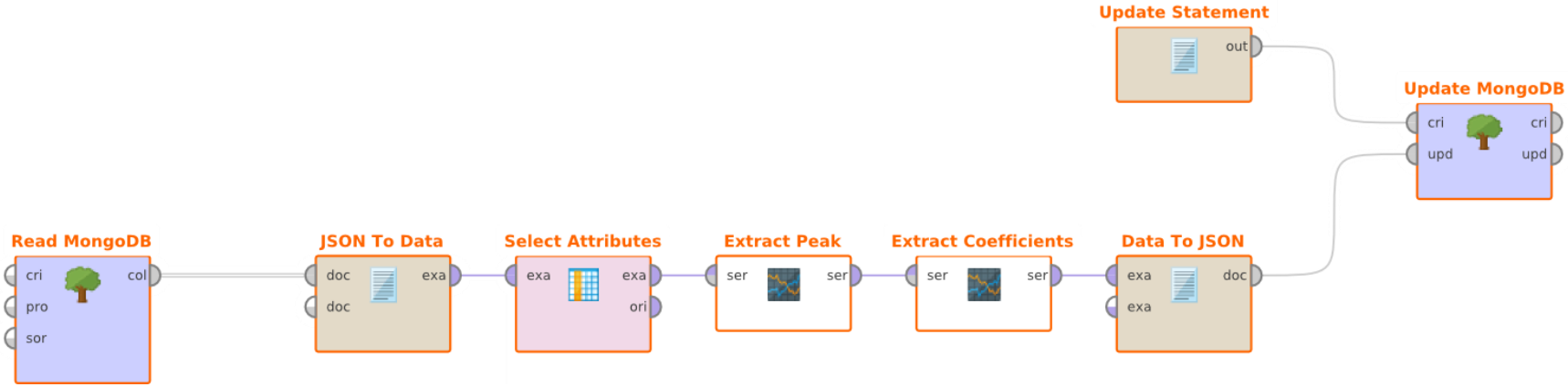
Data Enrichment



Data Enrichment



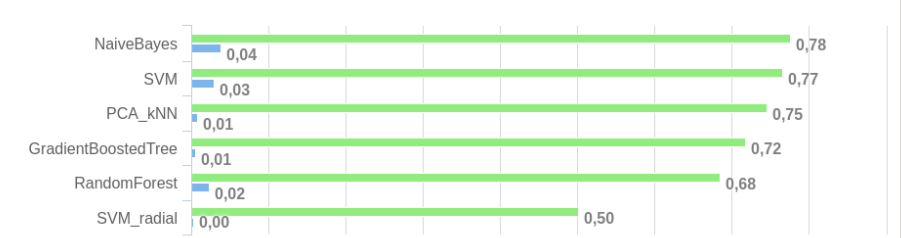
Data Enrichment in RapidMiner



RapidMiner Server

- Process Execution Core
  - REST API
  - Share data and processes
  - Schedule running processes
  - Dashboards
- Train, store and apply different machine learning techniques

Performances Graph



Methods Overview

MethodName	AUC	MongoDB Id (id)
NaiveBayes	0.776 +/- 0.037	58d9296af3dffd65acc35de3
SVM	0.766 +/- 0.030	58d93b71f3dffd65acc35de8
PCA_kNN	0.746 +/- 0.008	58d92995f3dffd65acc35de4
GradientBoostedTree	0.718 +/- 0.006	58d92d3df3dffd65acc35de7

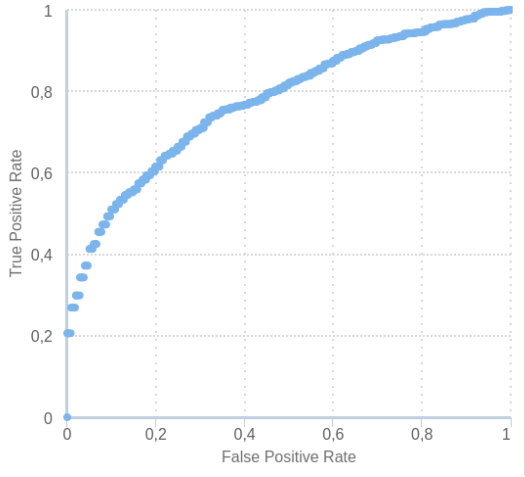
NaiveBayes: General Information

id	
CreationTime	2017.03.27 AD at 1
MethodName	NaiveBayes
Pathes.Model	/project/processes/(
Pathes.Parameter	/project/processes/(
Pathes.Performance	/project/processes/(
_id.$oid	58d9296af3dffd65a

NaiveBayes: Performance Values

Performance Criterion	Value	Standard Deviation	Varia
AUC	0.776	0.037	0.001
accuracy	0.710	0.031	0.001

NaiveBayes: ROC Curve



Conclusion

- Special requirements for processing steel production data
- Product oriented data architecture
  - Designed to handle raw production data
- Analytical platform to apply machine learning and data visualization

A REFERENCE ARCHITECTURE FOR QUALITY IMPROVEMENT IN STEEL PRODUCTION

rapidminer
www.rapidminer.com

Dr. Edwin Yaqub
Email: eyaqub@rapidminer.com
Web: rapidminer.com

David Arnu
Email: darnu@rapidminer.com
Web: rapidminer.com

RapidMiner GmbH
Westfalendamm 87
44141 Dortmund
Germany
+49 231 292 99 301

RapidMiner, Inc.
10 Fawcett St., 5th Floor
Cambridge MA 02138
United States
+1 617 401 7708