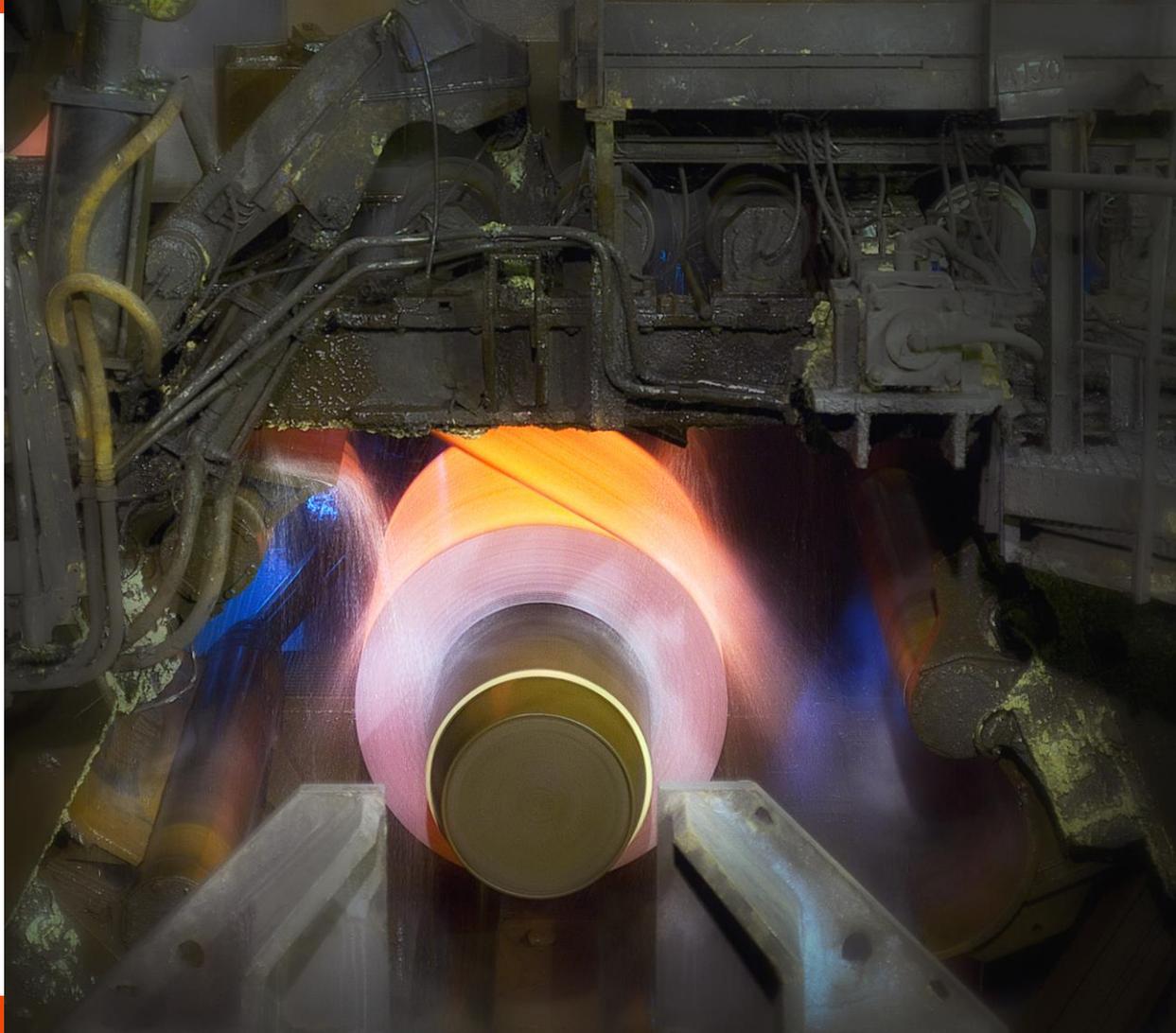


SMART DATA FÜR DIE QUALITÄTSKONTROLLE IN DER STAHLPRODUKTION

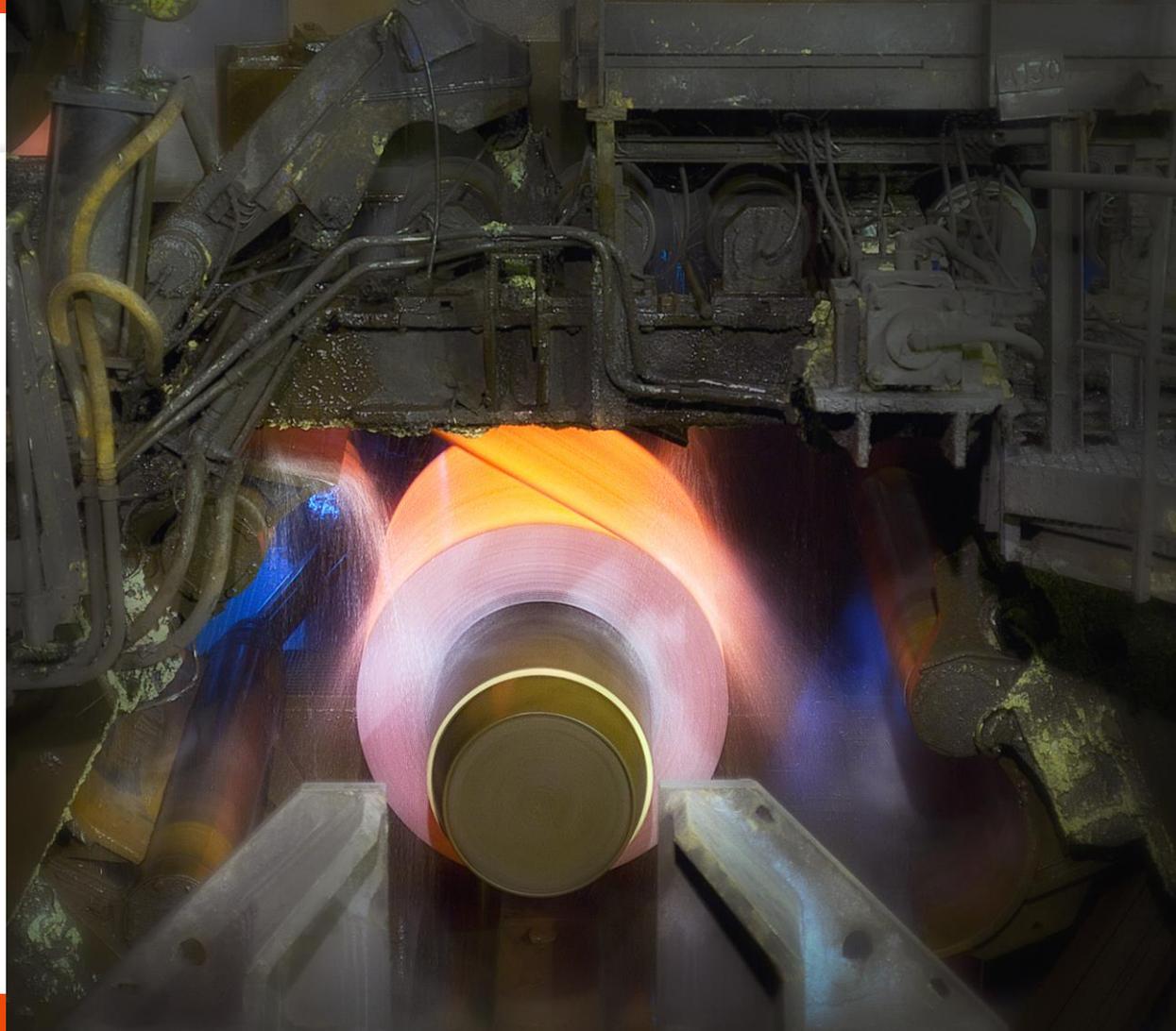
**David Arnu¹, Dr. Edwin Yaqub¹, Dr. Fabian Temme¹, Ralf
Klinkenberg¹, Dr. Marcus Neuer²**

¹ RapidMiner GmbH, ² BFI Forschungsinstitut

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

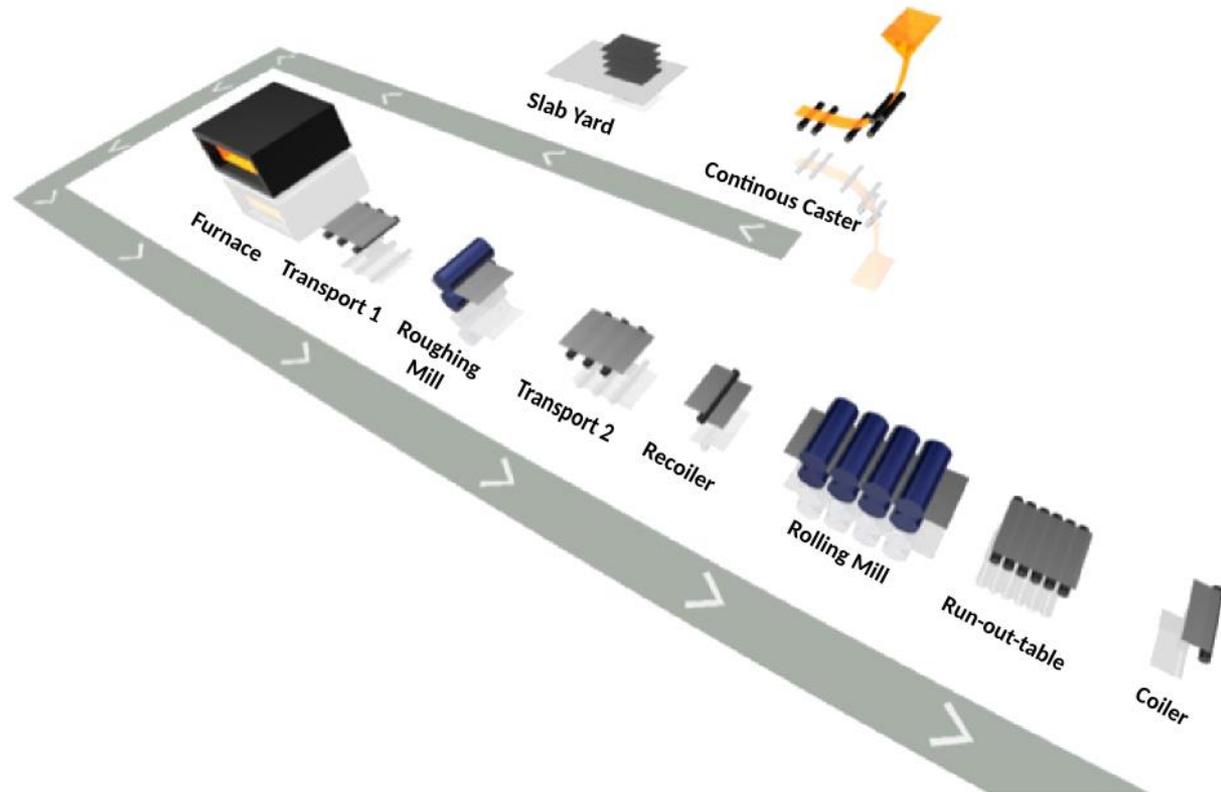


- Weltweite Nachfrage
- Kosten
 - Energie
 - Zeit
 - Ressourcen
- Datenlage
 - Hohe Variabilität
 - Große Datenmenge

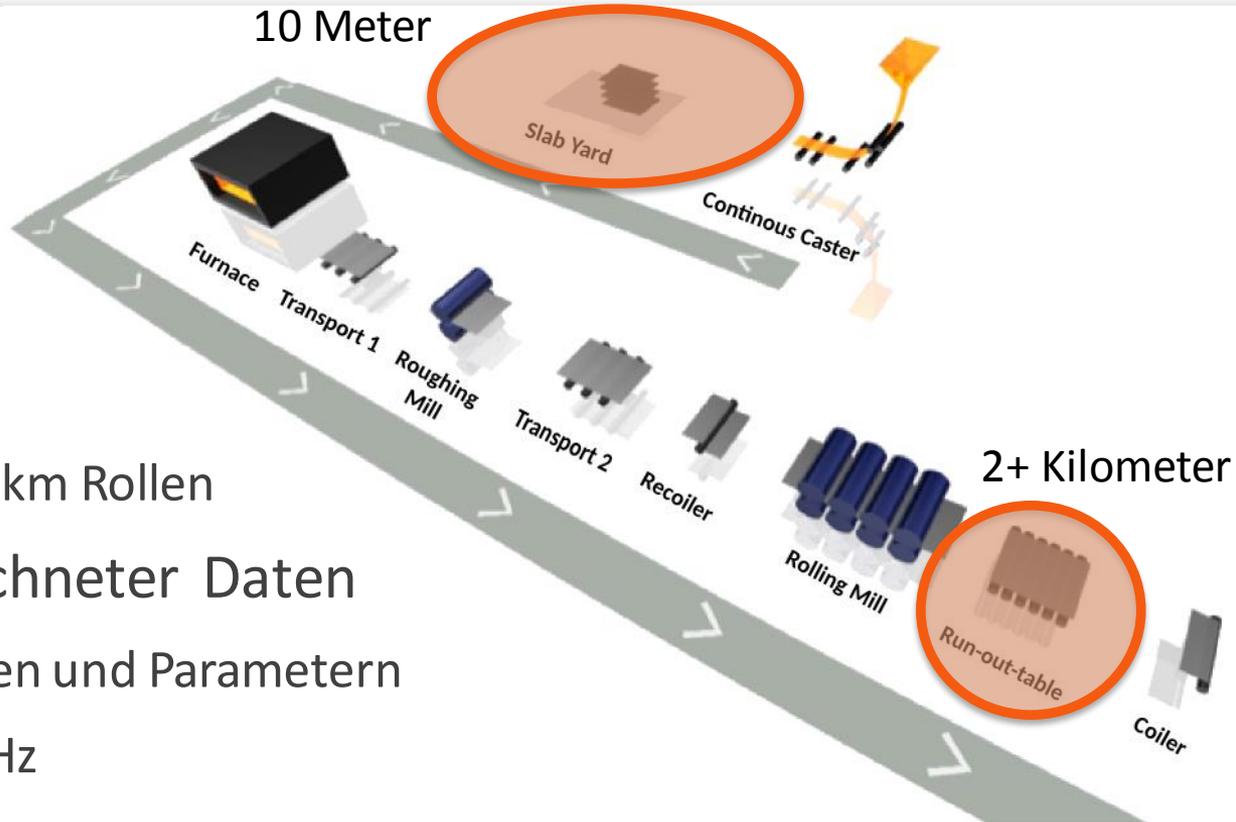


Fos-sur-Mer, France. All rights acquired provided photo credits are mentioned: Alain Sauvan/ArcelorMittal Fos-sur-Mer

Übersicht Stahlproduktion



Übersicht Stahlproduktion



- Vielseitiges Produkt
- Veränderliche Form
 - Von 10m Blöcken zu 2km Rollen
- Große Anzahl aufgezeichneter Daten
 - Hunderte von Sensoren und Parametern
 - Abtastraten von 1-10Hz
 - Produktionsdaten von mehreren Jahren

PRESED

Predictive Sensor Data Mining for Product Quality Improvement



Scuola Superiore
Sant'Anna



<http://www.pressed.eu/>

PRESED Ziele

- Verbesserte Qualitätsvorhersage
- Frühzeitige Erkennung von Produktdefekten
- Produktorientiertes Designkonzept
- Maschinelles Lernen für die Stahlproduktion
 - Informationsgewinn durch hochauflösende Produktionsdaten

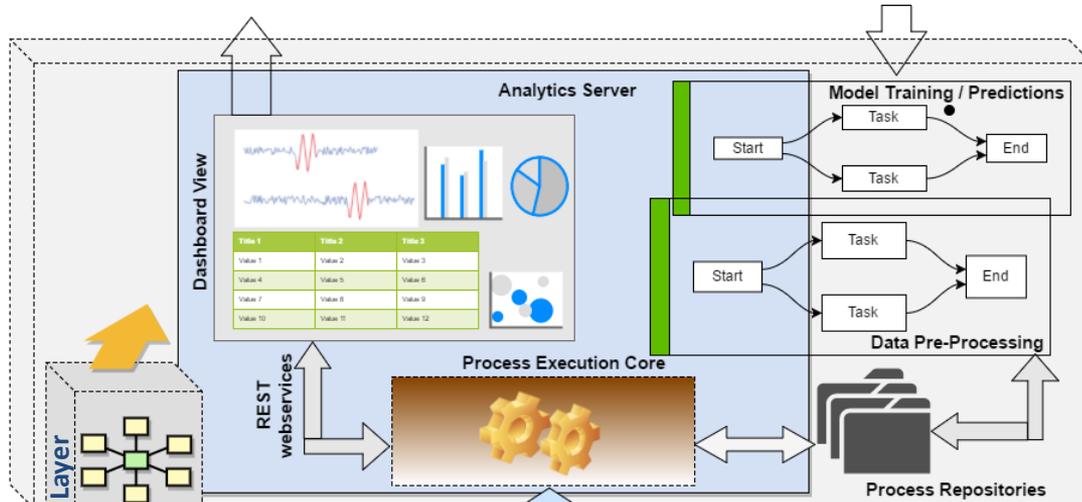


Referenzarchitektur, um diese Ziele zu erreichen



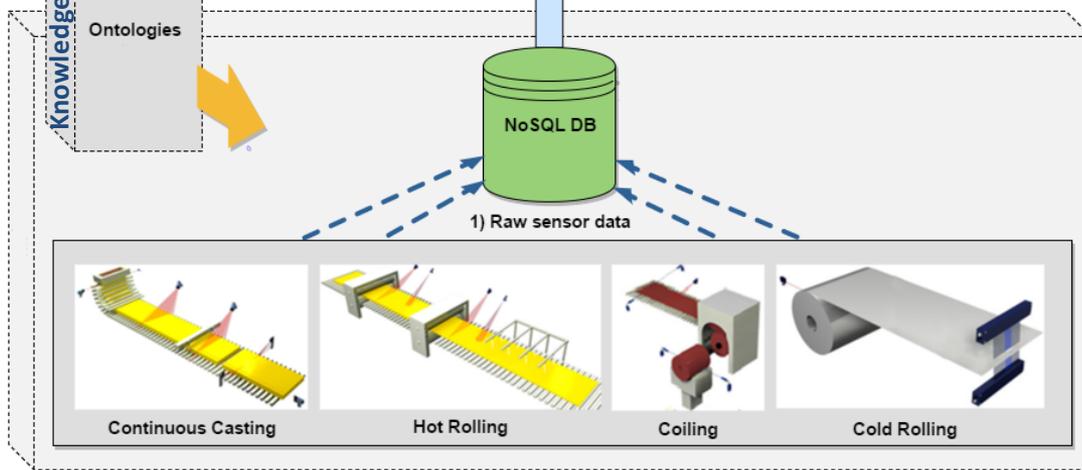
Plant Operators/
Manager

Analytical Layer



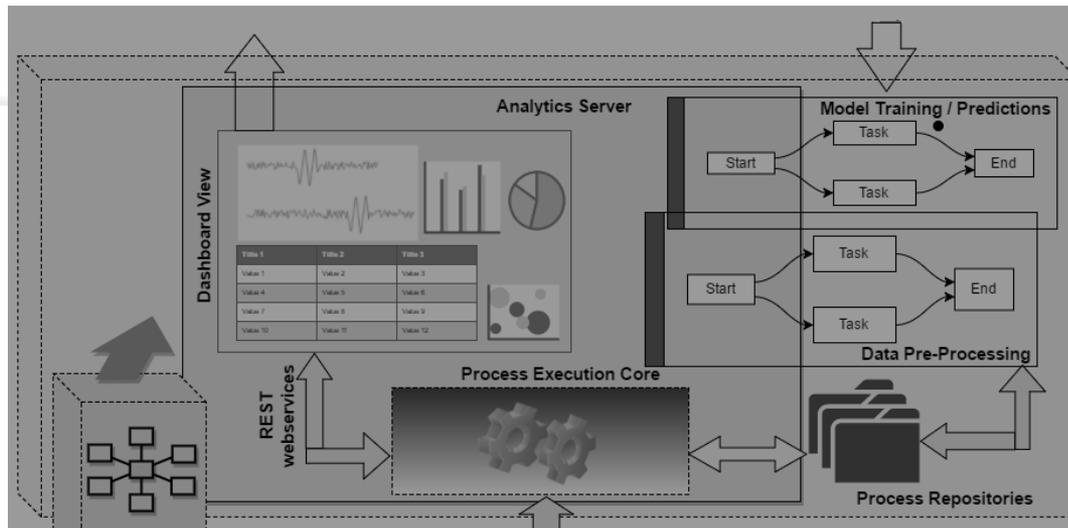
Domain Experts

Data Layer

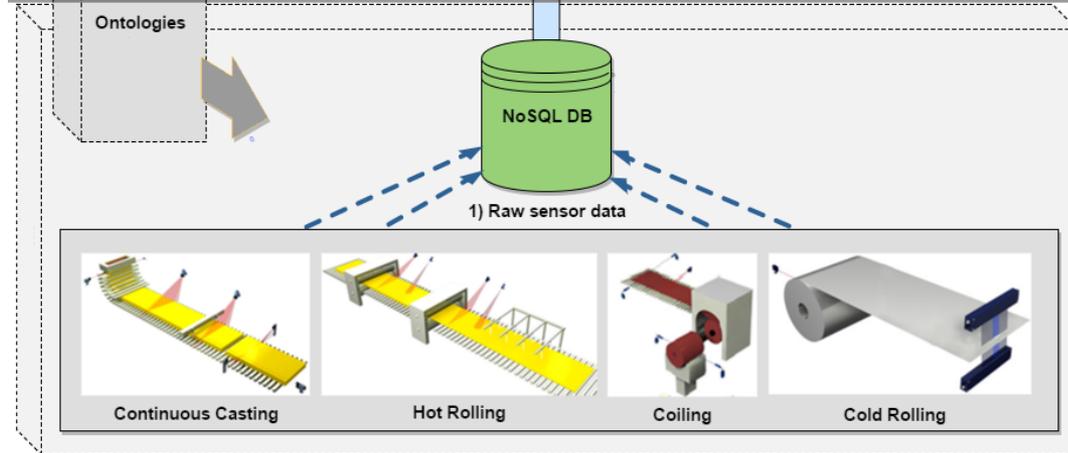


Data Engineers

Analytical Layer



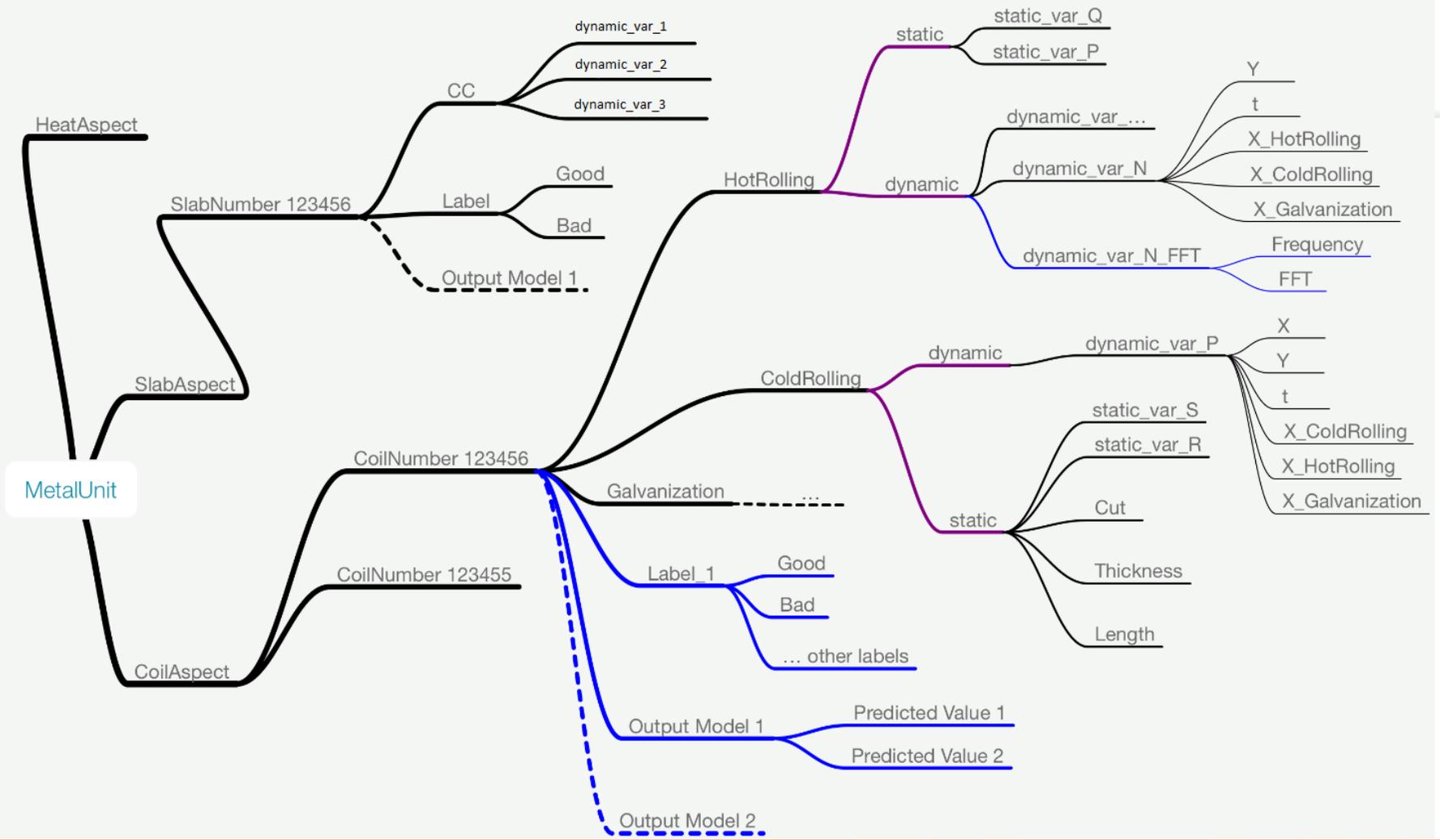
Data Layer



Data Engineers

Metal Unit (Datenschicht)

- Datenmodell für Sensordaten aus der Produktion
- Objektorientierte Struktur
 - Leicht veränderbare Datenstruktur
- Hierarchische Struktur, um verschiedene Produktionsschritte abzubilden

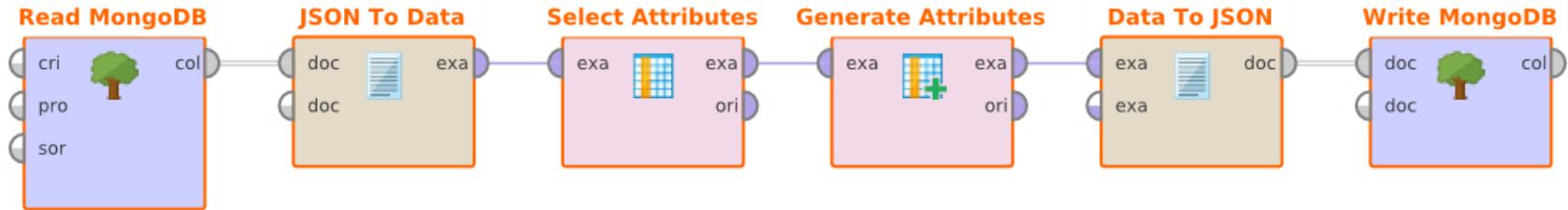


Werkzeuge (NoSQL)

- MongoDB Datenbank
 - OpenSource
 - Skalierbar
 - Vielseitige Abfragesprache
- Dokumentorientiert
- 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
  ]
  ...
  ...
}
```

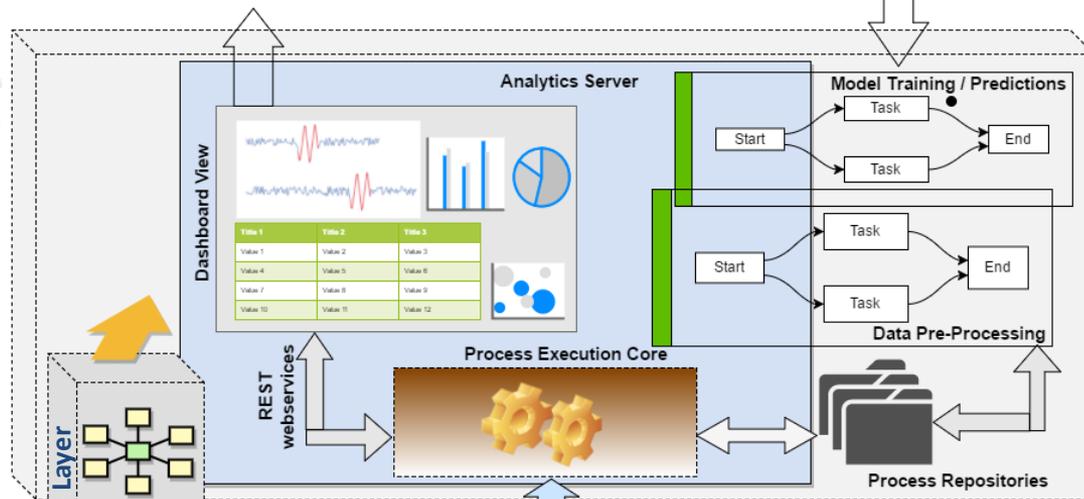
Werkzeuge (RapidMiner)





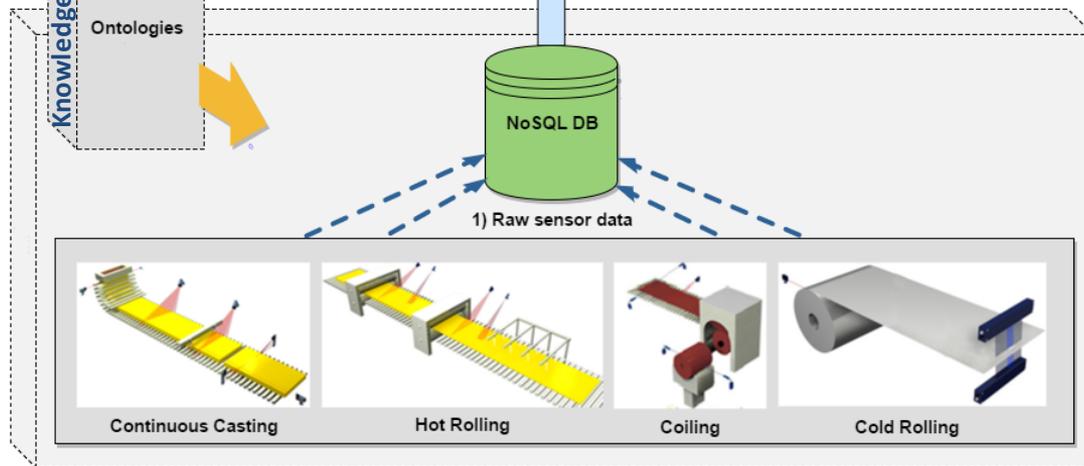
Plant Operators/
Manager

Analytical Layer



Domain Experts

Data Layer

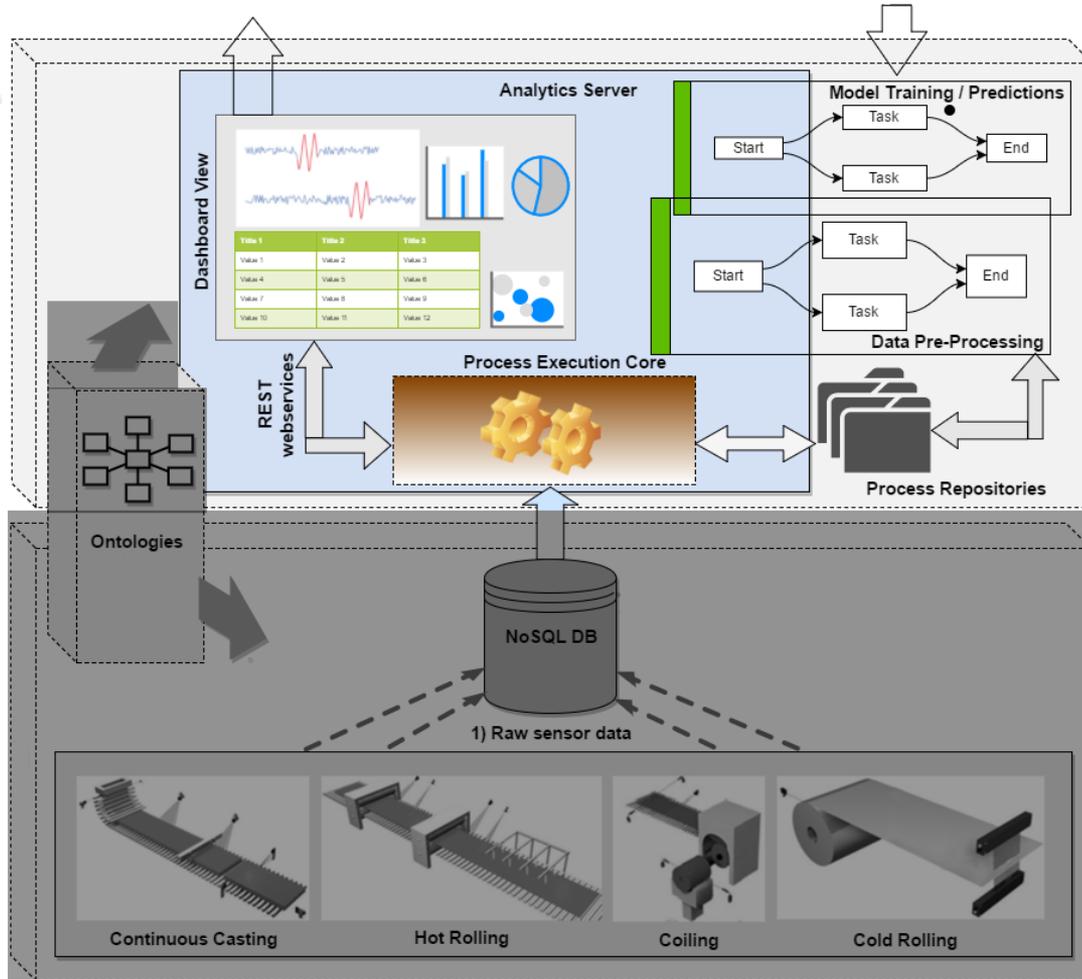


Data Engineers



Plant Operators/
Manager

Analytical Layer



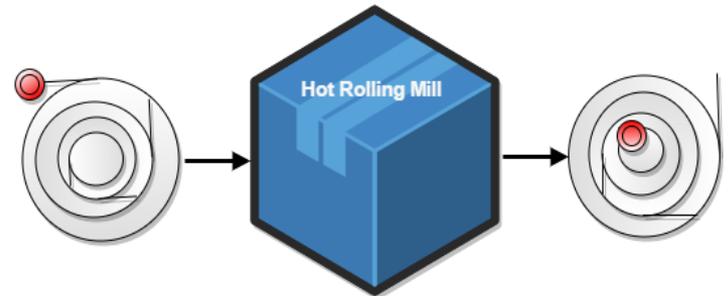
Domain Experts

Analyseschicht (Smart Data)

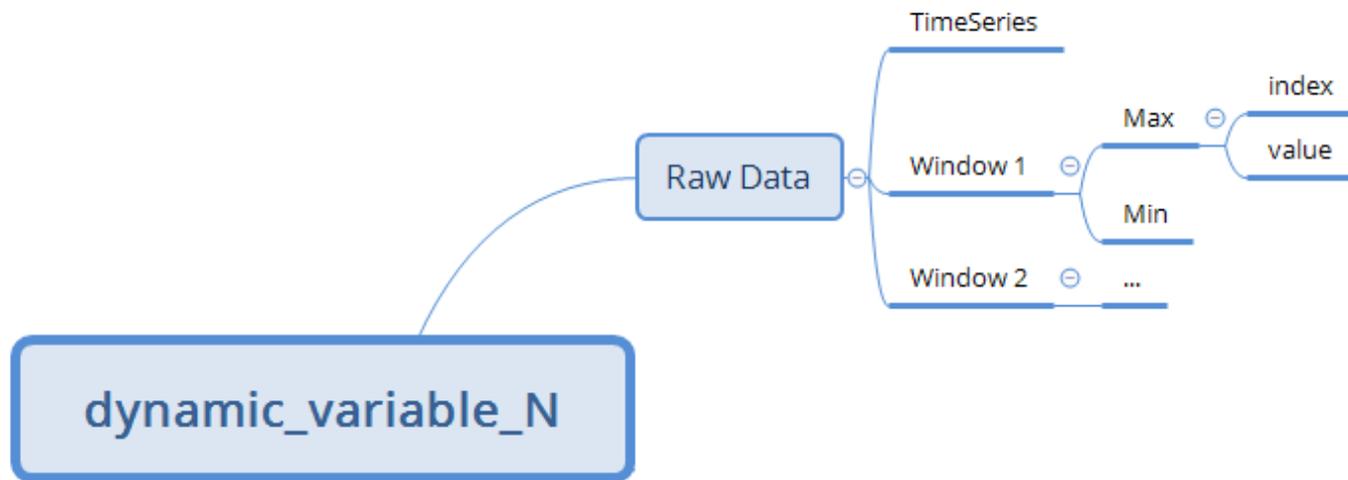
- *Smart Data* durch Datenanreicherung
 - Neue Signalrepräsentation (FFT, Filtern, Ableitung)
 - Merkmalsgenerierung (deskriptive Statistik, *Shapelets*)
- Skalierung der Daten
 - Nachverfolgung von Veränderungen während der Produktion
 - Veränderte Form des Produktes

Analyseschicht (Smart Data)

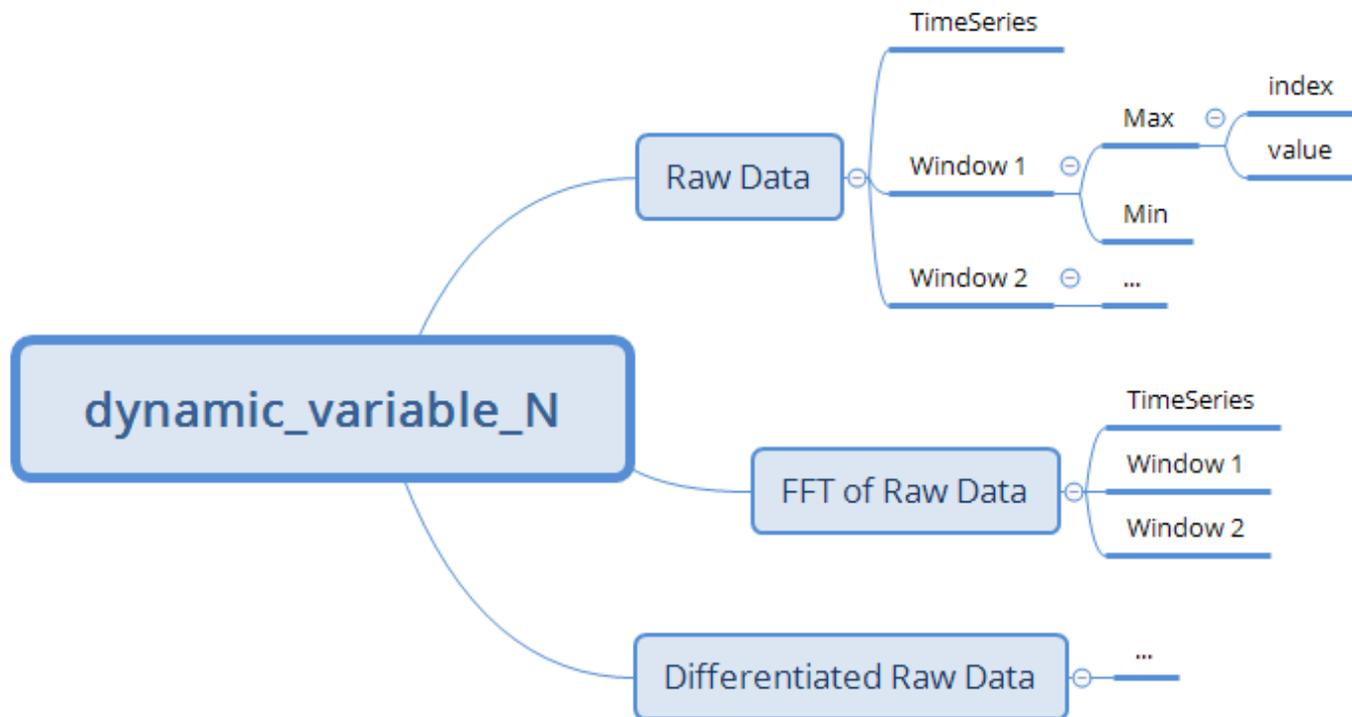
- *Smart Data* durch Datenanreicherung
 - Neue Signalrepräsentation (FFT, Filtern, Ableitung)
 - Merkmalsgenerierung (deskriptive Statistik, *Shapelets*)
- Skalierung der Daten
 - Nachverfolgung von Veränderungen während der Produktion
 - Veränderte Form des Produktes



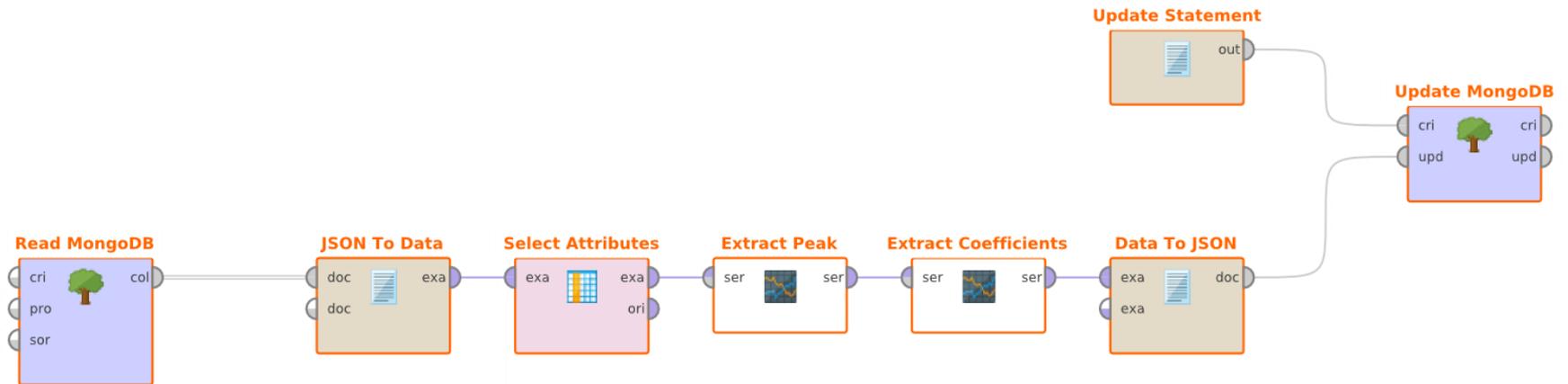
Datenanreicherung



Datenanreicherung



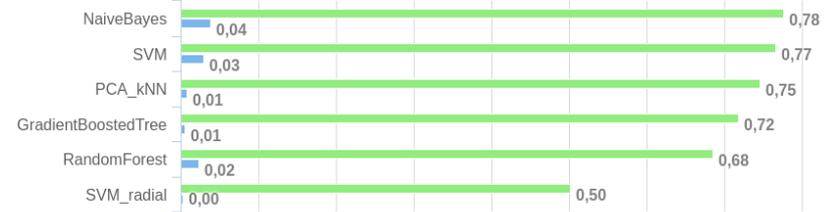
Datenanreicherung mit RapidMiner



RapidMiner Server

- Prozessanwendung
 - REST API
 - Daten- und Prozessverwaltung
 - Prozessausführung planen
 - Prozessvisualisierung
 - Benutzerverwaltung
- Trainieren, speichern und anwenden verschiedener Algorithmen

Performances Graph



Methods Overview

MethodName	AUC	MongoDB Id (id)
NaiveBayes	0.776 +/- 0.037	58d9296af3dff65acc35de3
SVM	0.766 +/- 0.030	58d93b71f3dff65acc35de8
PCA_kNN	0.746 +/- 0.008	58d92995f3dff65acc35de4
GradientBoostedTree	0.718 +/- 0.006	58d92d3df3dff65acc35de7
RandomForest	0.685 +/- 0.022	58d92b6f3dff65acc35de5

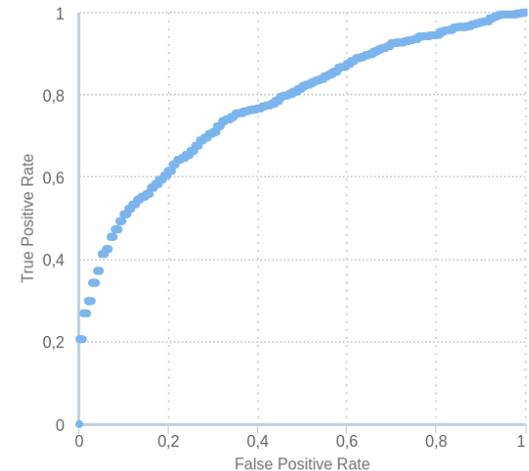
NaiveBayes: General Information

id	
CreationTime	2017.03.27 AD at 1
MethodName	NaiveBayes
Pathes.Model	/project/processes/
Pathes.Parameter	/project/processes/
Pathes.Performance	/project/processes/
_id.\$oid	58d9296af3dff65a

NaiveBayes: Performance Values

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

NaiveBayes: ROC Curve



Zusammenfassung

- Spezielle Anforderungen für Produktionsdaten der Stahlindustrie
- Produktorientiertes Datenmodell
 - Verarbeitung nicht komprimierter Produktionsdaten
- Analyseplattform für maschinelles Lernen und Datenvisualisierung



SMART DATA FÜR DIE QUALITÄTSKONTROLLE IN DER STAHLPRODUKTION



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