IDS 2017 – Industrial Data Science Conference

Dortmund, Germany | September 5th, 2017

Data Science Use Cases & Best Practices in Industry

Digitization, Industry 4.0, the Internet of Things (IoT), and the industrial internet are transforming industries leveraging big data in various forms such as streaming data, structured and unstructured data, text, image, audio, and sensor data using advanced analytics, posing significant challenges and offering enormous opportunities.

Join us at **IDS 2017,** to learn how advanced analytical applications are being used by world-leading organizations like ABB, Achenbach Buschhütten, Arcelor Mittal, BMW, Daimler, DEW, Lufthansa, Miele, Volkswagen, and others to get the most value out of their data, which challenges they encountered, how they solved them, and which tools they used to get a competitive advantage.

Twitter: #ids2017

Web: http://ids2017.rapidminer.com/

E-Mail: ids2017@rapidminer.com

rapidminer

tu technische universität dortmund

SFB 876 Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung

IPS Institute of Production Systems

CPS.HUB /NRW
Competence Center for Cyber Physical Systems

IDS 2017 – Industrial Data Science Conferece

Dortmund, Germany | September 5th, 2017

Welcome to
IDS 2017

Data Science Use Cases & Best Practices in Industry

Ralf Klinkenberg
Co-Founder & Head of Data Science
RapidMiner

Twitter: #ids2017

Web: http://ids2017.rapidminer.com/

E-Mail: ids2017@rapidminer.com

IDS 2017 Conference Chairs

Ralf Klinkenberg

Co-Founder & Head of Data Science Research

RapidMiner

Prof. Dr.-Ing. Jochen Deuse

Head of the Institute of Production Systems (IPS)

TU Dortmund University

Dr. Stefan Michaelis

General Manager

Collaborative Research Center SFB 876

TU Dortmund University

IDS 2017 Programme & Organization Committee

- **Prof. Dr. -Ing. Jochen Deuse,** Institute of Production Systems (IPS), TU Dortmund University

- **Prof. Dr. Katharina Morik,** Head of Artificial Intelligence Group & Speaker of Collaborative Research Center SFB 876, TU Dortmund Univ.

- **Dr. Stefan Michaelis,** General Manager of Collaborative Research Center SFB 876, TU Dortmund University

- **Ralf Klinkenberg,** Co-Founder & Head of Data Science Research, RapidMiner

- **Dr. Martin Schmitz**, Head of Data Science Services, RapidMiner

- **Julian Schallow,** General Manager, IPS Engineers GmbH

- **Monika Gatzke**, CPS.Hub NRW

- **Jacqueline Schmitt,** Institute of Production Systems (IPS), TU Dortmund University

- **Anika Altmann,** Event Management, RapidMiner

- **Jennifer Paulsen**, Event Management & Marketing, RapidMiner

- **Mario Wiegand,** Institute for Research and Transfer (RIF), TU Dortmund University

- **Dr. Fabian Temme,** Data Scientist, RapidMiner

- **David Arnu,** Senior Data Scientist, RapidMiner

- **Edin Klapic,** Data Scientist, RapidMiner

- **Dr. Edwin Yaqub,** Data Scientist, RapidMiner
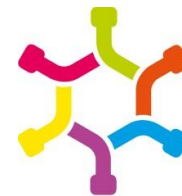
- **Philipp Schlunder**, Data Scientist, RapidMiner

IDS 2017 Supporting Organizations

**Welcome to
IDS 2017
and
Overview of
Industrial
Data Science
Use Cases**

Ralf Klinkenberg

www.rapidminer.com

rklinkenberg@rapidminer.com

Gartner *predicts that by 2018, more than half of large organizations globally will compete using advanced analytics and proprietary algorithms, causing the disruption of entire industries.*
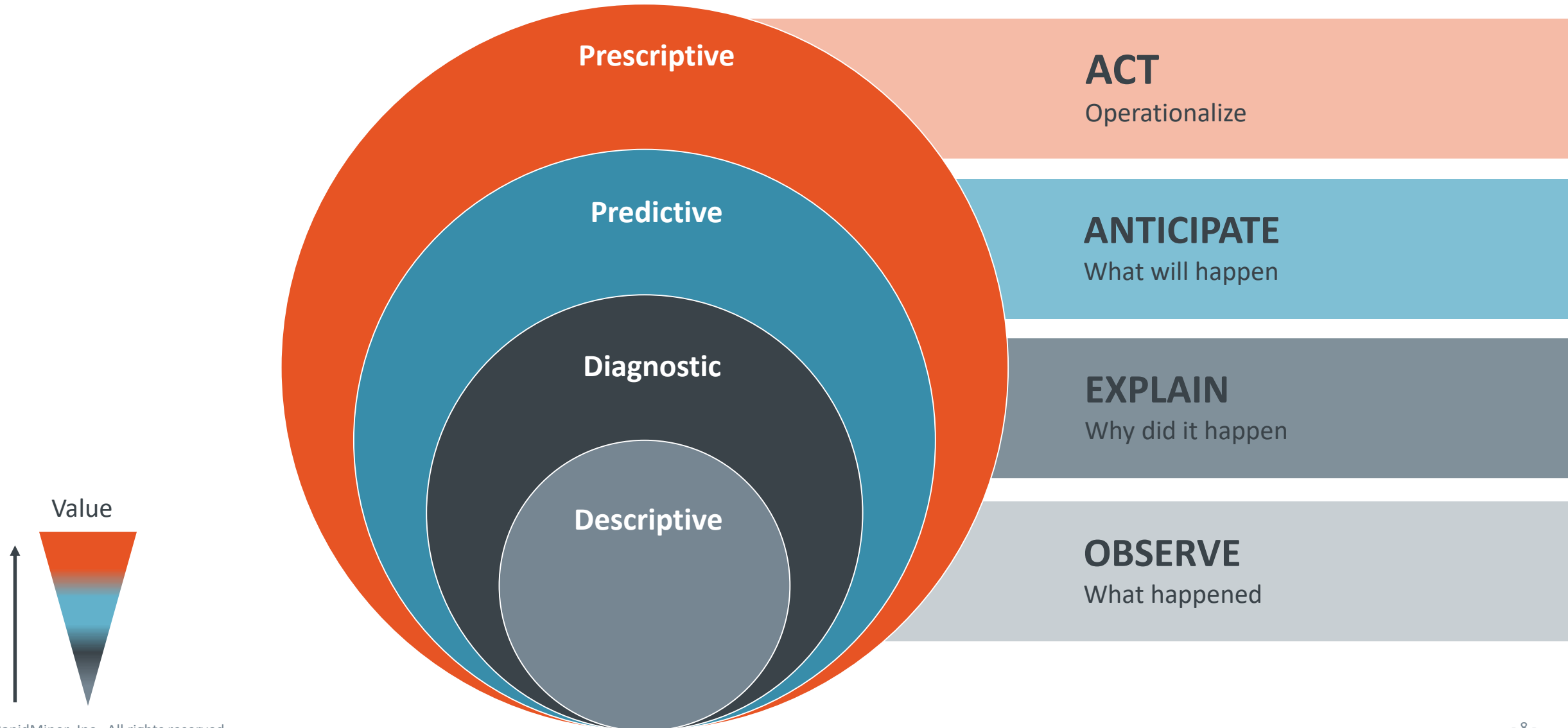
Real data science, fast and simple.

Predictive Analytics Will **DISRUPT** Markets

rapidminer

"Gartner predicts that by 2018, more than half of large organizations globally will **compete using advanced analytics and proprietary algorithms,** causing the disruption of entire industries."

Gartner.

Predictive Analytics is **GAME CHANGING**

rapidminer

ANTICIPATE TOMORROW

- Make **HIGHLY ACCURATE** predictions, forecasts or classifications

- **OPERATIONALIZE** in mission-critical systems to drive decisions and actions in near real time
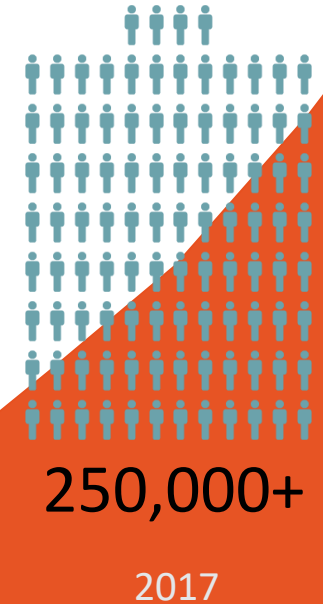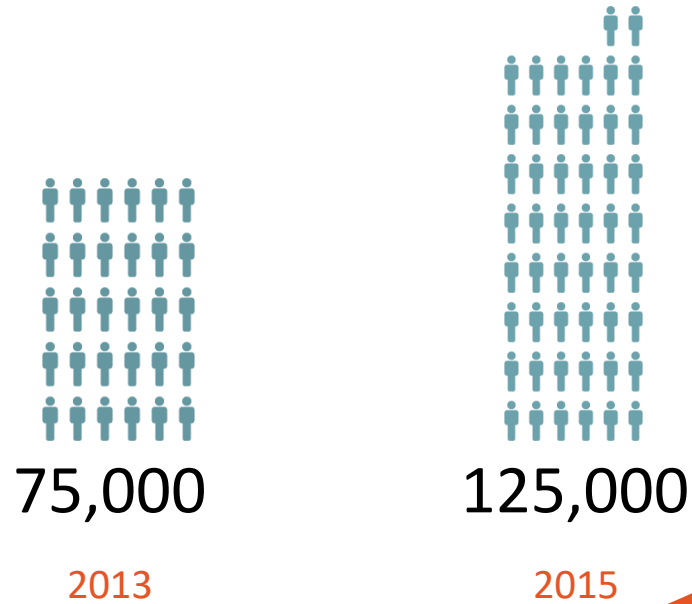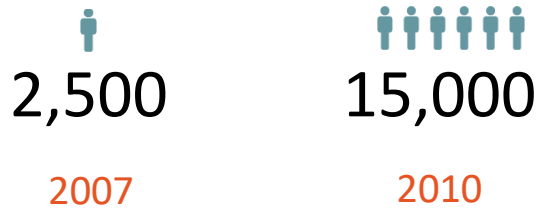
#1 Open Source Community

EMERGING **EXPANDING** **INTENSIFYING**

2,500 15,000 75,000 125,000 250,000+

2007 2010 2013 2015 2017

Number of Registered Users

http://community.rapidminer.com/

RapidMiner **Highlights**

By the numbers

#1	250,000+	250+	50+
Open Source Data Science Platform	Engaged Community Members	Global Clients	Channel Partners

Analysts

Gartner
Leader
2014, 2015, 2016 & 2017
Gartner Magic Quadrant for Advanced Analytics Platforms

FORRESTER®
Strong Performer
2015 & 2017
Forrester Wave on Big Data Predictive Analytics 2015 /Data Science Platforms 2017

KDnuggets
#1 Open-Source Platform
Last five years in a row
Data Mining & Analytics Software Poll

Howard Dresner Dresner Advisory Services, LLC
Innovation Winner
2015
Wisdom of Crowds for Advanced & Predictive Analytics, Big Data Analytics & End-User Data Prep

Accolades

AI 100
CB Insights
The AI 100, 2017
"100 Startups Using Artificial Intelligence to Transform Industries"

VENTANA RESEARCH 2016 Technology Innovation **AWARDS**
VENTANA RESEARCH
2016 Technology Innovation Awards Winner
Predictive Analytics

Data Science Use Cases

Sample Customer Use **Cases**

rapidminer

Voice of the Customer

Automated Customer Feedback Text Analysis for Automated E-Mail / Social Media, Categorization, Triage & Routing

Manufacturing – Production Optimization

Optimization Of Production Logistics & Flows, Quality, Yield, Product Mix, Process Mining

Manufacturing – Predictive Maintenance

High Value Assets - Silicon, Cars, Trucks, Aircraft, Turbines, IT Infrastructure,...

Fraud Detection

Fraud detection in retail network historical data on service usage, transaction history, customer profiles, usage logs, and known cases of fraudulent behavior

Maximizing Customer Lifetime Value

CRM applications including optimization of direct marketing campaigns, automated generation of product recommendations for cross-selling and up-selling, customer churn prevention, and fraud detection

Predictive Maintenance

Customers Using RapidMiner for Predictive Maintenance, i.e. for Predicting & Preventing Machine Failures before they happen:

- Major German Car Manufacturer:
  Text Analytics of Repair & Service Reports to Identify
  Car Quality & Car Maintenance Issues

- Major European & South American Airplane Manufacturers
  and Major International Airplane Operators:
  Sensor Data & Text Mining Repair & Service Reports for
  Predictive Airplane Maintenance & Resource Allocation

- Major European Cement Producer:
  Cement Mill Failure Prediction & Prevention

- Major Chinese Energy Provider:
  Wind Turbine Failure Prediction & Prevention

Predictive Maintenance: Cement Mills

One of the largest two cement producers world-wide:

- RapidMiner-based solution for predicting and preventing drilling machine failure

- RapidMiner-based solution for simulating drilling machine behavior when changing system parameters

- 40 persons trained to deploy the RapidMiner-based solution in their cement mills world-wide
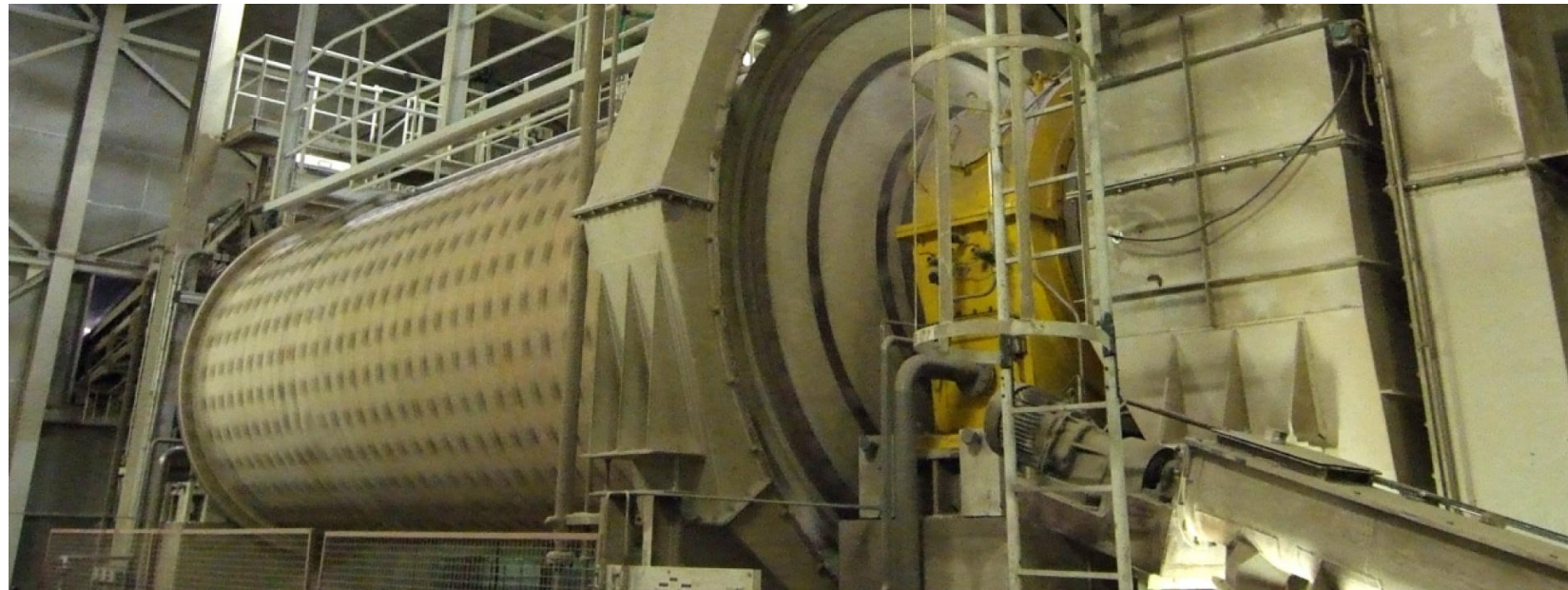
OPTIMAL MIXTURE OF INGREDIENTS?

Optimizing Mixtures of Ingridients

- Which mixtures will produce high quality products?
- Which mixtures will lead to quality issues?
- How much of particular expensive additives is needed?
- How to lower costs while ensuring high product quality?
- How to increase production process reliability & product quality?
- What variables are correlated to product quality and how?
- How to predict and ensure product quality?
- How much of each ingredient is optimal?
- How to configure the production process and machines?
- => Automated Predictions & Alerts & Action Recommendations
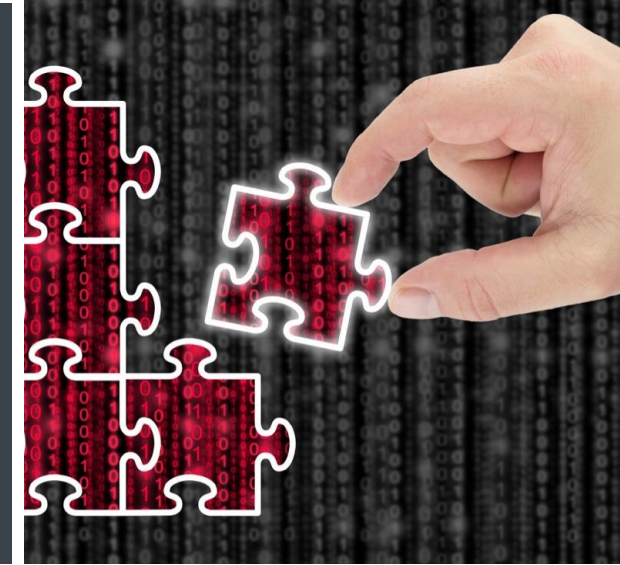- => Lower Cost & Lower Risk & Higher Reliability & Higher Quality

RapidMiner Predictive Analytics Platform

The RapidMiner Competitive Advantage

rapidminer

Unified Platform

Prototype – Substantiate – Operationalize – seamless, high performance orchestration

Lightning Fast Machine Learning

Powerful, visual & guided use of 1,500 data prep and ML functions & third party ML libraries

#1 Marketplace for Data Science Expertise

On-demand consultants, algorithms & extensions; global presence & domain expertise in every industry

Real data science, fast and simple.

RapidMiner Studio

All-In-One Data Science Workflow Designer
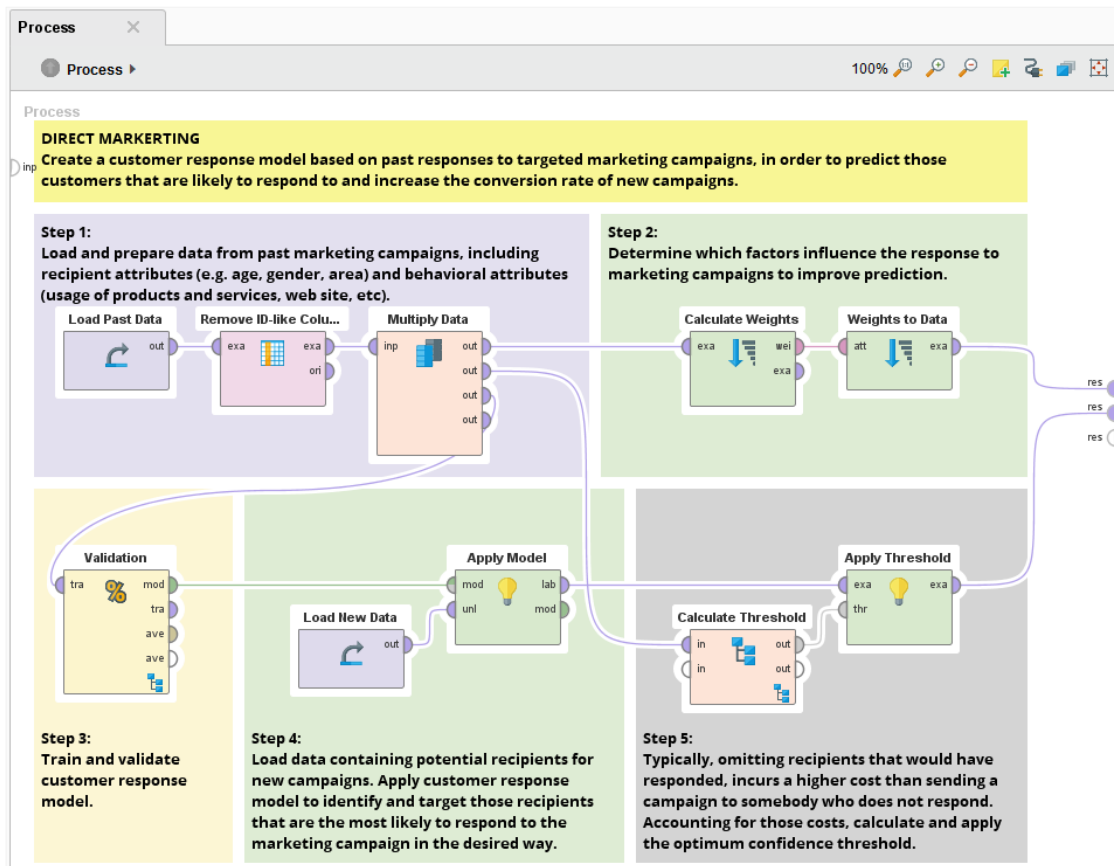
Lightning Fast
Visual interface for rapidly building complete analytic workflows

Powerful
Rich library of algorithms and functions to build the strongest possible model for any use case

Open & Extensible
Open source innovation keeps pace with changing business needs

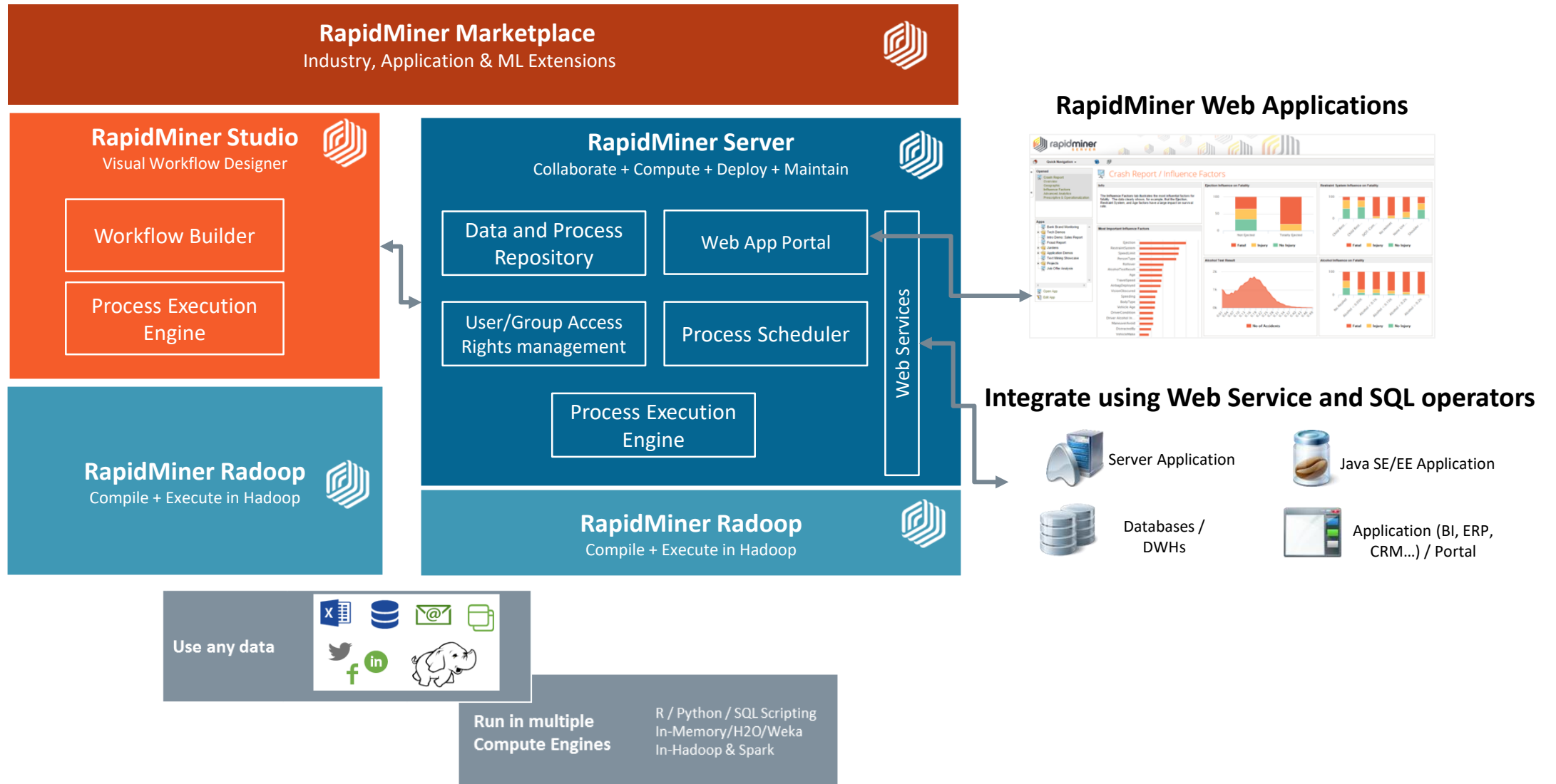# The RapidMiner **Platform**

RapidMiner Marketplace
Industry, Application & ML Extensions

RapidMiner Studio
Visual Workflow Designer

Workflow Builder

Process Execution Engine

RapidMiner Radoop
Compile + Execute in Hadoop

RapidMiner Server
Collaborate + Compute + Deploy + Maintain

Data and Process Repository

Web App Portal

User/Group Access Rights management

Process Scheduler

Process Execution Engine

Web Services

RapidMiner Radoop
Compile + Execute in Hadoop

RapidMiner Web Applications

Integrate using Web Service and SQL operators

Server Application

Java SE/EE Application

Databases / DWHs

Application (BI, ERP, CRM...) / Portal

Use any data

Run in multiple Compute Engines

R / Python / SQL Scripting
In-Memory/H2O/Weka
In-Hadoop & Spark

Current State & Challenges & Barrierers for Data Science in Industry

Current State of Industrial Data Science

- Insular solutions – only few and isolated use cases

- Data silos – data not sufficiently used across the organization

- Experts of different fields rarely communicate enough

- Almost no truely completely data-driven companies
  - but Google and Tesla create pressure on automotive sector
  - but Google NEST creates pressure on utility providers
  - but high risk of others gaining a significant competitive advantage

- The opportunities are enormous, but most companies do not dare to move ahead fast enough but wait for others
  - Listen closely today – You will see others are already moving ahead!

Challenges & Barrierers for IDS

- Different mind sets of domain experts and data analysts

- Lack of communication and lack of mutual understanding

- Lack of ideas for use cases and potential opportunities

- Risk-averseness and lack of investment

- Lack of openness for innovation

- Lack of leadership and management support

=> IDS 2017 wants to inspire and show opportunities and successful examples to make you move ahead faster!

CRISP-DM
CRoss-Industry Standard Process for Data Mining

CRISP-DM – The Standard for DM Projects
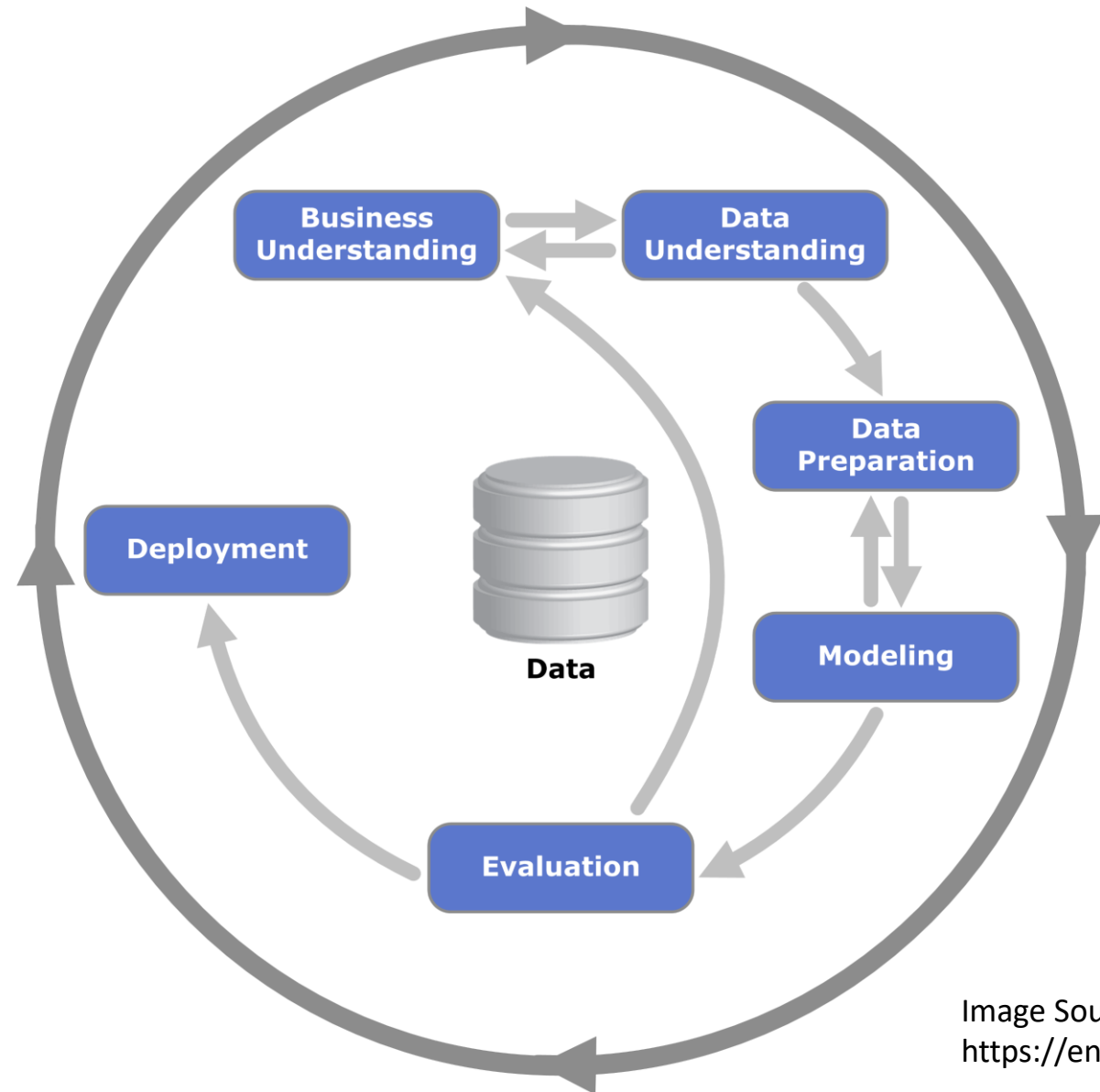
Data

CRoss-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining:

- Structured and proven process for a close cooperation of domain experts, data experts, and data scientists
- Cooperative iterative process with communication during all phases of the process fostering mutual understanding
- Best combination of domain exprtise and human knowledge with data-driven machine learning based models
- Goal-, validation-, and deployment-oriented

Image Source: Wikipedia:
https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

Industrial Data Science
Use Cases & Best Practices
Presented at IDS 2017

IDS 2017 Theme following CRISP-DM

- Thought Leaders – from Research to Industry Applications
- Data & Data Analysis Platforms, Approaches, Use Cases
- Data Preprocessing & Transformation and Use Cases
- Modeling & Prediction and Use Cases

- … all presented along industry use cases & best practices
- … intermitted by breaks & networking opportunities

Prof. Dr. Katharina Morik

Machine Learning and Data Science: Research and Applications in Industry

- Head of Artificial Intelligence Unit, TU Dortmund University
- Speaker of the Collaborative Research Center SFB 876
- Member of AcaTech
- Internationally Acknowledged Expert for Machine Learning
- Initiator of Machine Learning Research in Germany
- Broad Experience in Algorithmic Research as well as Industry Applications

Prof. Dr.-Ing. Jochen Deuse

From Industrial Engineering towards Industrial Data Science

- Head of Institute of Production Systems (IPS), TU Dortmund University
- Held senior management positions in the Bosch Group in Germany and Australia
- Head of the chair of Industrial Engineering since 2005, which in 2012 merged with the chair of Industrial Robotics and Production Automation to form the Institute of Production Systems (IPS) under his direction.
- Member of the board at the industry network NIRO e.V.
- Expert for Industrial Engineering and Data Science Applications in Industry

11:00-11:30h Coffee Break & Networking

From Prototype to Operative Software – Data Analytics @ Lufthansa

Dr. Fabian Werner

Data Science Consultant

Lufthansa Industry Solutions

Optilink – A Big Data Platform for Industrial Cloud Applications

Roger Feist

Head of Automation

Achenbach Buschhütten

Dynamic Bottleneck Forecasting in Flexible Manufacturing Systems

Ferdinand Klenner

Project Leader Predictive Analytics and Optimization Flexible Manufacturing Systems

BMW

13:00-14:00h Lunch & Networking

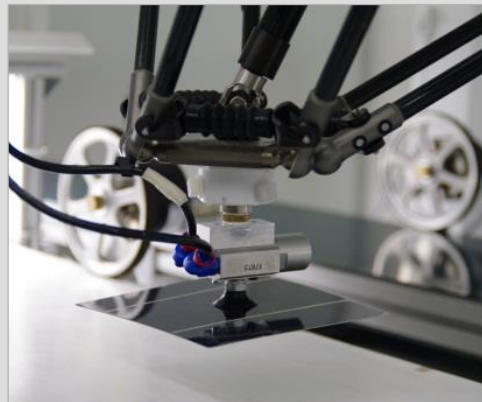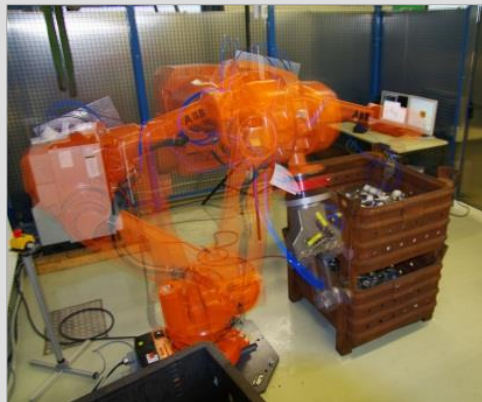# Predicting Assembly Times & Assembly Plans for New Product Designs & Variants

Project "Pro Mondi – Predicting Assembly Plans in Digital Factories" sponsored by the German Government

Application of Data Mining for Prospective Assembly Time Determination

Dr. Olga Erohin

Corporate Development Division Professional Technology

Miele

Ralf Kretschmer

Director Segment Professional Laundry Technology

Miele

Data-Mining-Based Generation of Assembly Work Plans to Accelerate Product Emergence Processes

Dr. Regina Wallis

Strategic Series Management

Daimler

Prediction Model for Agile Rework Scheduling

Sven Krzoska

Expert for Industrial Engineering and Data Mining

Volkswagen

15:30-16:00h Coffee Break & Networking

EVERYTHING OK?

BIG DATA CAN BE OVERWHELMING

Holistic Development of Industrial Big-Data Applications and Services
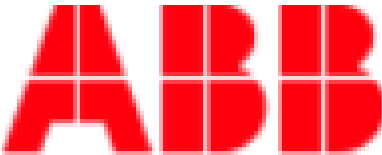
Marcel Dix

Scientist

Industrial

Data Analytics

ABB

Dr. Benjamin Klöpper
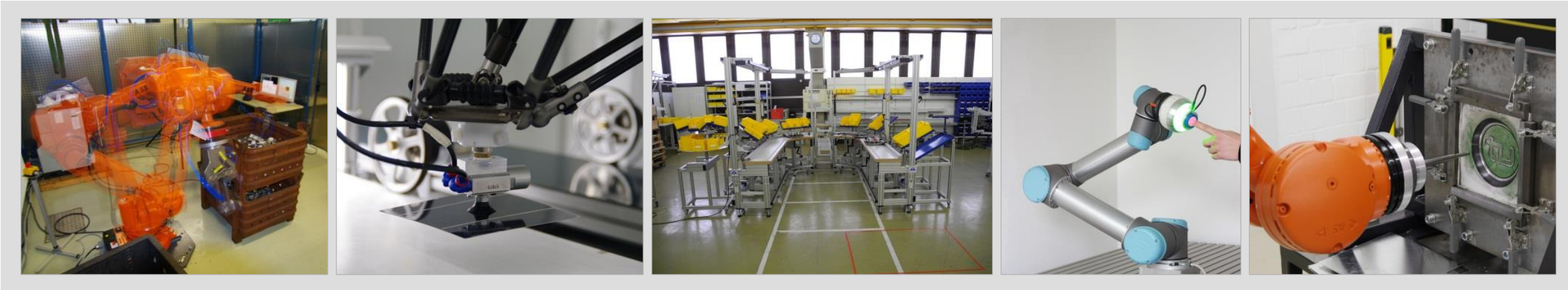
Principal Scientist

Industrial

Data Analytics

ABB

Detecting, Predicting, and Preventing Exceeded Emissions & Critical Situations

Project "FEE" sponsored by the German Government

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Quality Prediction in a Rolling Mill from Sensor Data Streams

Daniel Lieber

Deputy Operations Manager Rolling Mill and Forging Shop Witten

DEW - Deutsche Edelstahlwerke Specialty Steel GmbH & Co. KG

17:30-19:30h Networking Reception