

Industry Applications of Machine Learning and Data Science

Ralf Klinkenberg, Co-Founder & Head of Data Science Research, **RapidMiner**
rklinkenberg@rapidminer.com
www.RapidMiner.com

Industrial Data Science Conference (IDS 2019), Dortmund, Germany, March 13th, 2019



Can you predict
the future?

A high-speed photograph of water being poured from a glass, creating a large, dynamic splash. The water is captured in mid-air, with many droplets and a large, curved sheet of water. The background is a light, neutral color.

Predictive Analytics finds the hidden patterns in big data and uses them to predict future events.

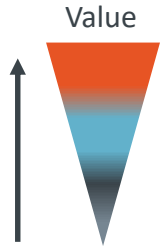
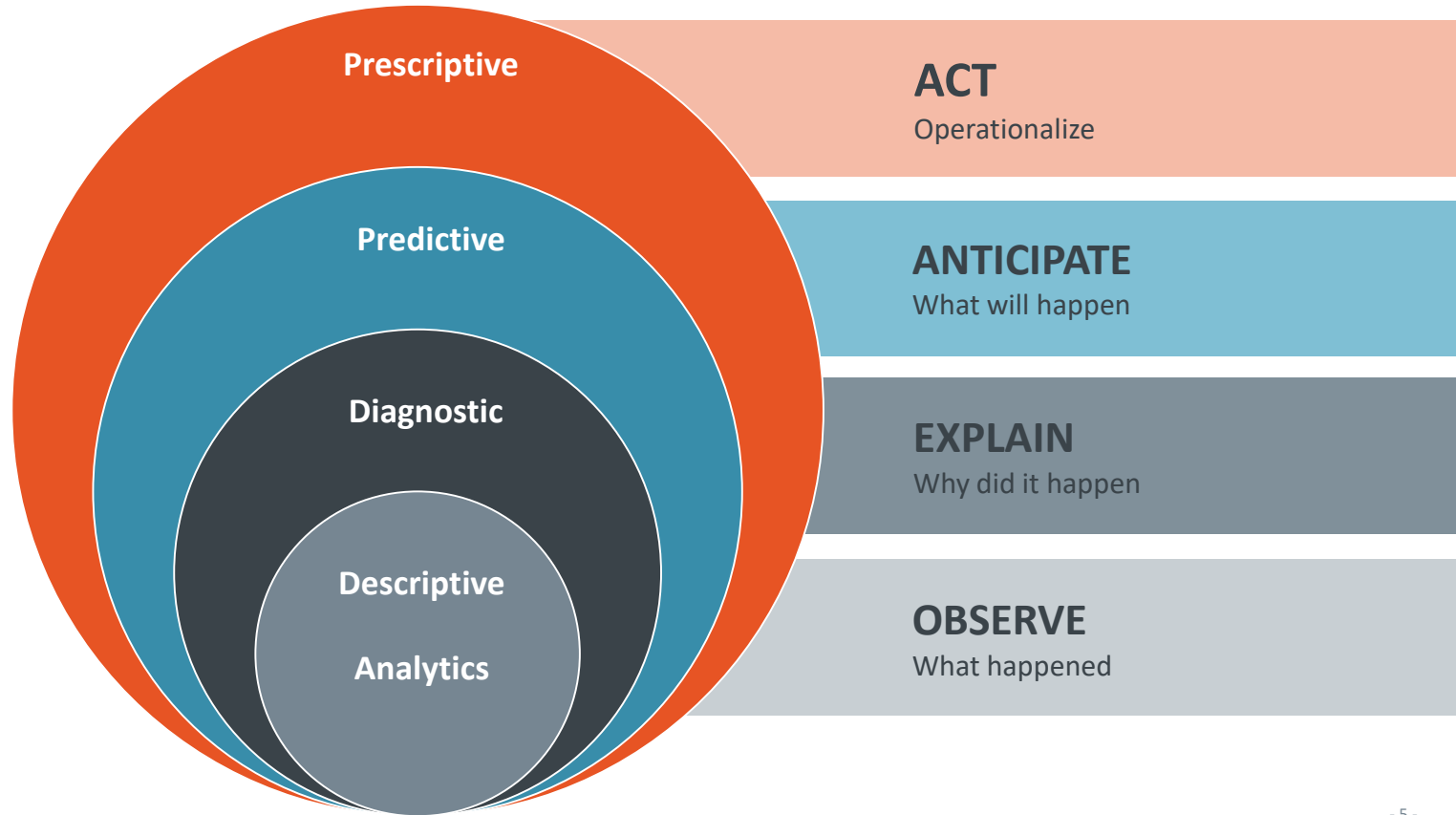
473ms

Patterns?

Machine Learning:

Pattern Detection,
Trend Detection,
Finding Correlations
& Causal Relations,
etc. from Data to
Create Models
that Enable the
Automated
Classification of
New Cases,
Forecasting of Events
or Values, Prediction
of Risks
& Opportunities

Predictive Analytics Transforms Insight into **ACTION**





Industry Applications of Machine Learning and Predictive Analytics

The background of the slide is a photograph of a large crowd at a concert or festival. The scene is illuminated by bright, warm stage lights, creating a hazy, golden atmosphere. Many people in the crowd have their arms raised, with some making 'rock on' hand gestures. The text '17 Million Customers (E.ON)' is overlaid on the upper part of the image.

17 Million
Customers (E.ON)

Really loyal

About to churn

Customer Churn Prediction:

Predict Which Customers are about to Churn.

Energy Provider E.ON: 17 Million Customers.

- => Predict & Prevent Churn
- => Secure Revenue, Less Costly Than Acquiring New Customers.

33 Million

Transactions per Year (Libri)

Will become a bestseller



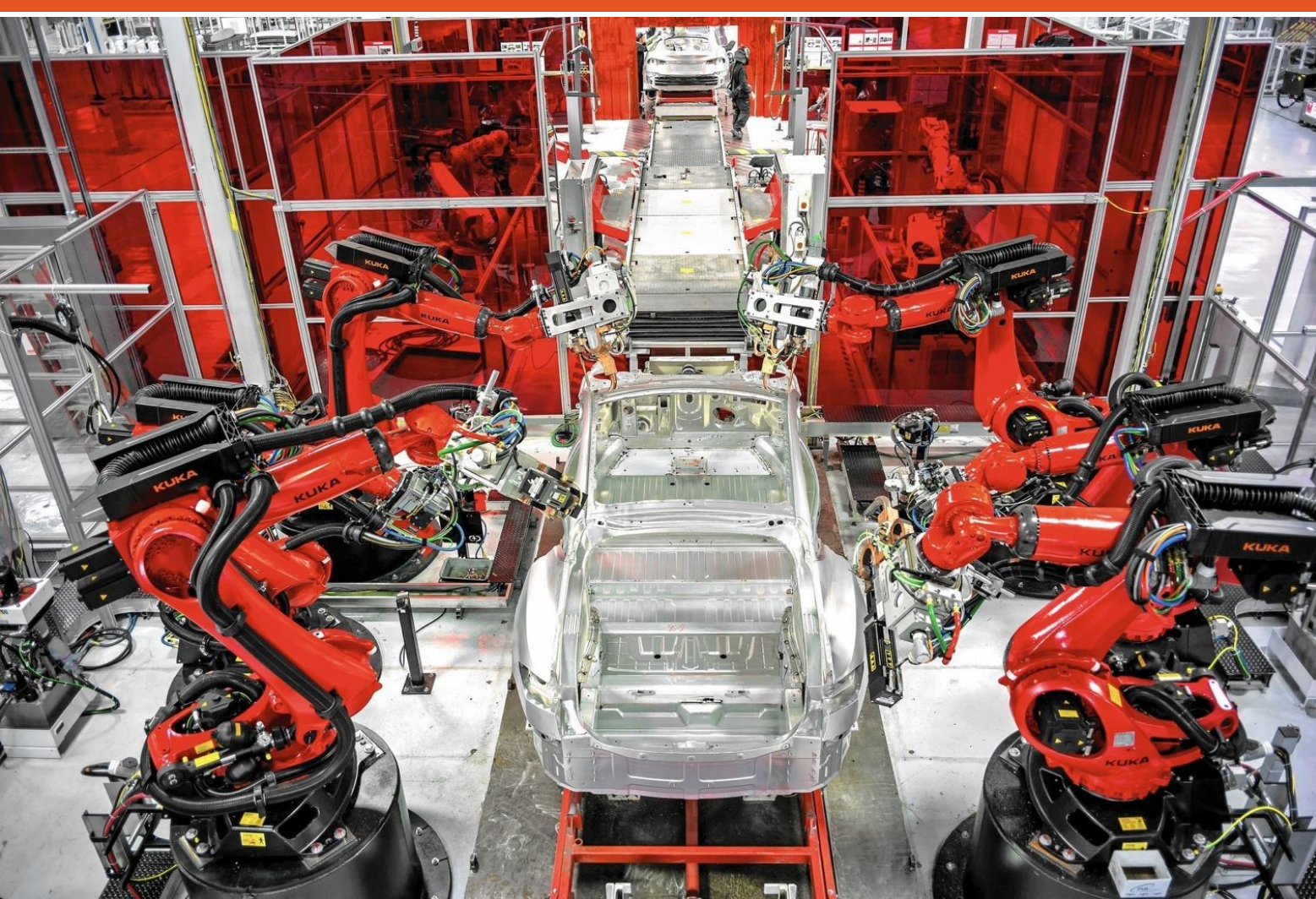
Demand Forecasting:

Predict Which
Book Will be Sold
How Often in
Which Region.

Book Retailer
Libri:

33 Million
Transactions
per Year.

=> Guarantee
Availability and
Delivery Times.



Predictive Maintenance:

Predict Machine Failures before They Happen in order to Prevent Them,

=>

Demand-Based Maintenance,
Fewer Failures,
Lower Costs

Predictive Maintenance





Sensor Data

Log Data

Meta Data

Free Text

Audio/Image Data

Predictive Maintenance

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

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



Engine Quality Prediction

- **Task:**
 - Predict Engine Lifetime / Quality from Audio Data
- **Challenges:**
 - Audio Feature Generation
- **Solution:**
 - Automated Feature Generation & Selection
 - Automated Optimizers
 - Classification Problem

Product Surface Quality Optimization

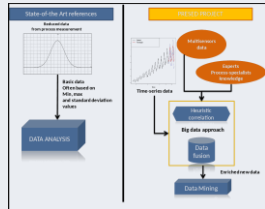
- **Task:**
 - Image Analysis to Determine Product Surface Quality/Issues
- **Challenges:**
 - Image Feature Construction
 - Light Conditions and Their Variety
- **Solution:**
 - Automated Feature Generation & Selection
 - Automated Optimization
 - Classification Problem

Project Overview

- Improve production quality in steel mills
- Optimize the manufacturing process by identifying the main causes of bad quality
- Predict the quality of the product as soon as possible to better characterize it and reduce the cost

Use Case

- Identifying causes for bad quality
- Predict the quality of a product as soon as possible
- Integrate expert knowledge for different production sites



Approach

- Build a Big Data Infrastructure for product tracking and visualization
- Extract features from raw sensor data series
- Apply machine learning for quality prediction
- Build an analytics server for model construction and management
- Enrich data with knowledge from experts

Challenges

- Design a data structure to track the material over the complete production process
- Tracking is difficult as the product changes shape (rolling, cutting)
- Huge amount of raw data (several hundred parameters, with a frequency of 1-10Hz over a 2-3 year period)

Contact: research@rapidminer.com www.rapidminer.com

Project partners:



Funded by:

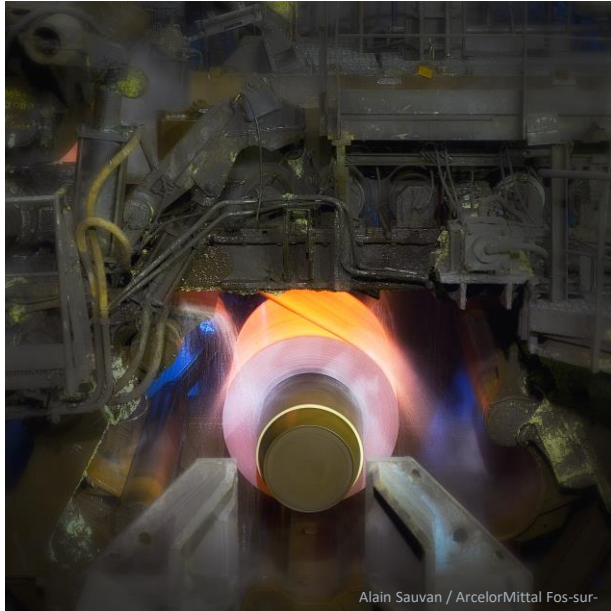


RFSR-CT-2014-00031

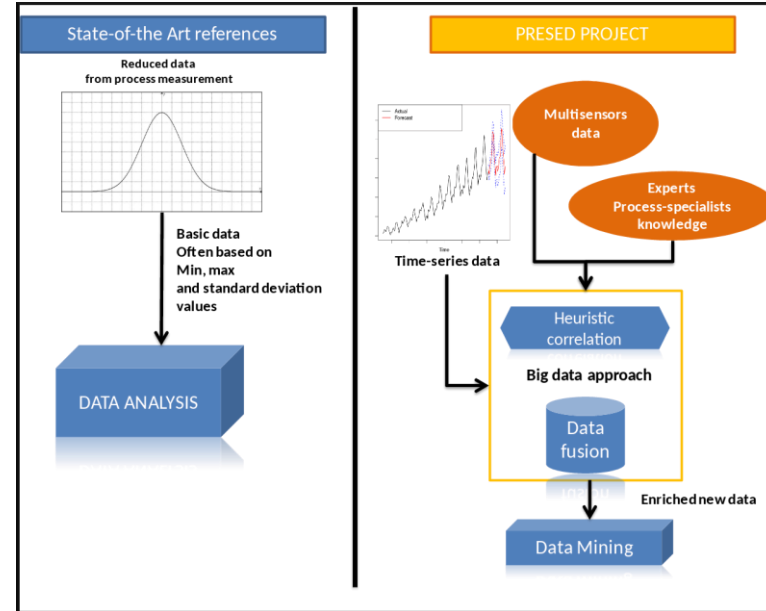
Project Overview

- Develop new methodologies to improve the quality of the steel production by
 - Identifying causes for bad quality
 - Predict the quality of a product as soon as possible during the production process
- RapidMiner building and providing the analytic server infrastructure
- RapidMiner implementing the tools developed by the other partners
 - New algorithms for time series analysis
 - Visualization methods for provided data

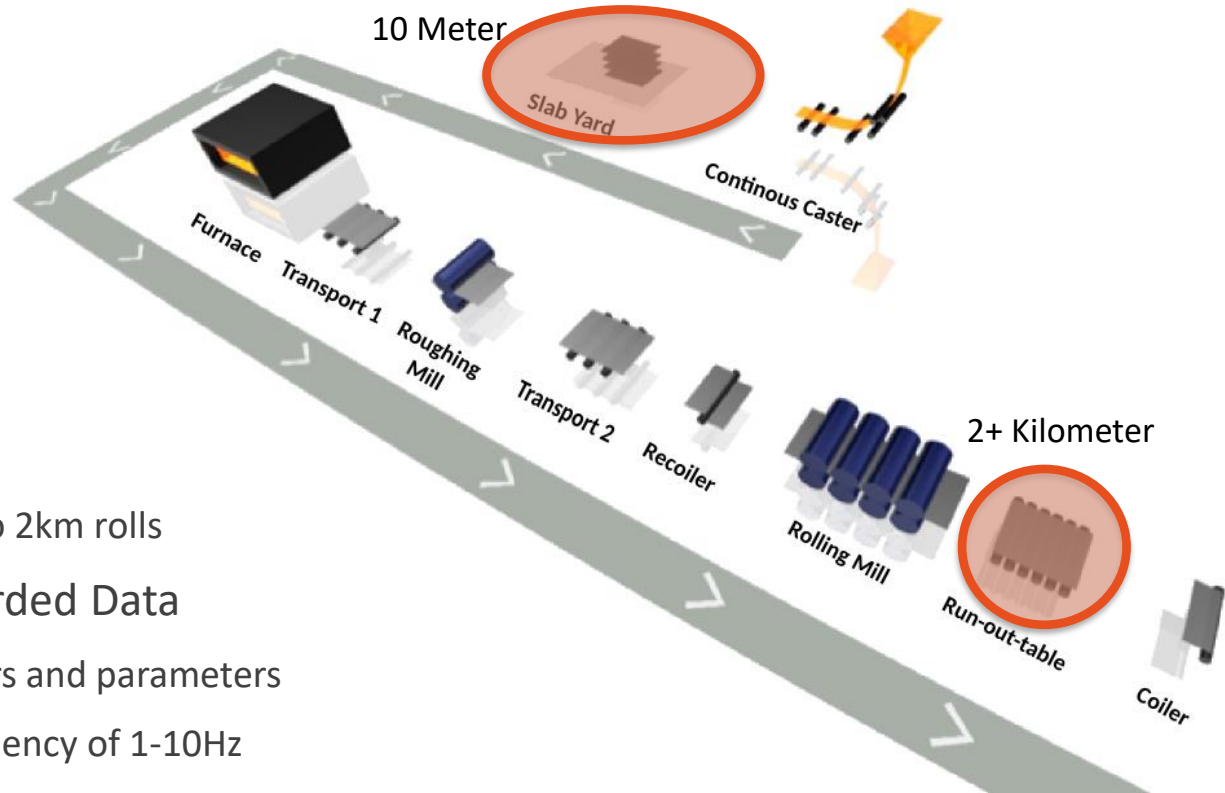
Product Quality Prediction with Machine Learning on Time Series Data



Alain Sauvan / ArcelorMittal Fos-sur-Mer



PRESED



- Product Variations
- Varying Form/Shape
 - From 10m blocks to 2km rolls
- Large Volume of Recorded Data
 - Hundreds of sensors and parameters
 - Value sensing frequency of 1-10Hz
 - Production data of many years



OPTIMAL MIXTURE OF INGREDIENTS?

Optimizing Mixtures of Ingredients

- 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

Optimizing Mixtures of Ingredients

Customers Using RapidMiner for Optimizing Mixtures of Ingredients:

- Major European Tire Manufacturer:
Optimizing the rubber ingredients to optimize product quality and features (e.g. durability, adhesiveness, etc.)
- Major European Metal Product/Part Producer
(Components of Cars, Trains, and Appliances):
Predicting & Optimizing Quality & Cost & Reliability
(ensure required quality level while reducing cost of ingredients, e.g. reducing expensive additives as much as reasonable, but not beyond)

EVERYTHING OK?





**BIG DATA CAN BE
OVERWHELMING**



Detecting & Predicting Critical Situations



- Big Data can be overwhelming – huge amounts of structured and unstructured data (e.g. texts) from many different sources
- How to find the relevant information to look at?
- How to effectively detect critical situations in a mass of data?
- How to predict and prevent critical situations?
- How to schedule maintenance early enough in advance?
- => Automated Alerts & Action Recommendations
- => Predicting and Preventing Machine Failures,
Predictive and Preventive Maintenance
- => Resource Optimization & Planning
- => Lower Cost & Lower Risk & Higher Reliability & Higher Quality

Detecting, Predicting, and Preventing Exceeded Emissions & Critical Situations



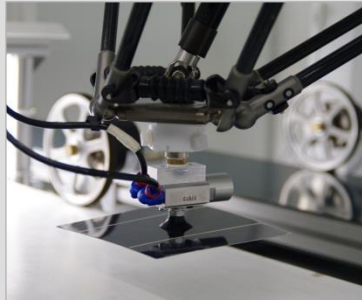
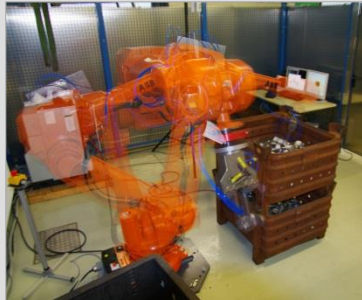
Project “FEE”
sponsored by the
German
Government



GEFÖRDERT VOM

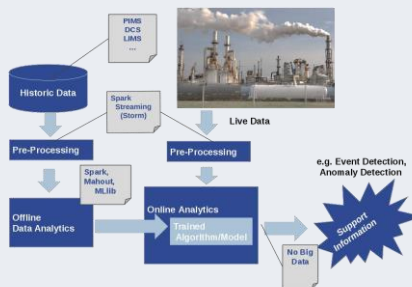


Bundesministerium
für Bildung
und Forschung



Project Overview

- Chemical plants are highly automated production sites that produce a lot of data
- Data comes from various sources (Sensor values, alarm logs, shift books)
- Critical operation conditions can result in a cascade of warnings (alarm shower)



Goal of the Project

- Design and build a Big Data infrastructure to manage the data of the production plants
- Develop methods for early detection of critical situations
- Develop methods to help users during critical situations
- Build Ad-hoc analysis functions to build intervention strategies
- Change from reactive to proactive action

Proposed Solution

- Usage of long-term data collections (~ 10 years) from production
- Offline data analytics with Big Data technologies
- Fast and adaptive online methods to evaluate current situations
- User-centric interaction concepts

Contact: research@rapidminer.com www.rapidminer.com

Project partners:

RapidMiner GmbH
ABB AG (Consortium Manager)
Technische Universität Dresden
Universität Kassel

Application partners:

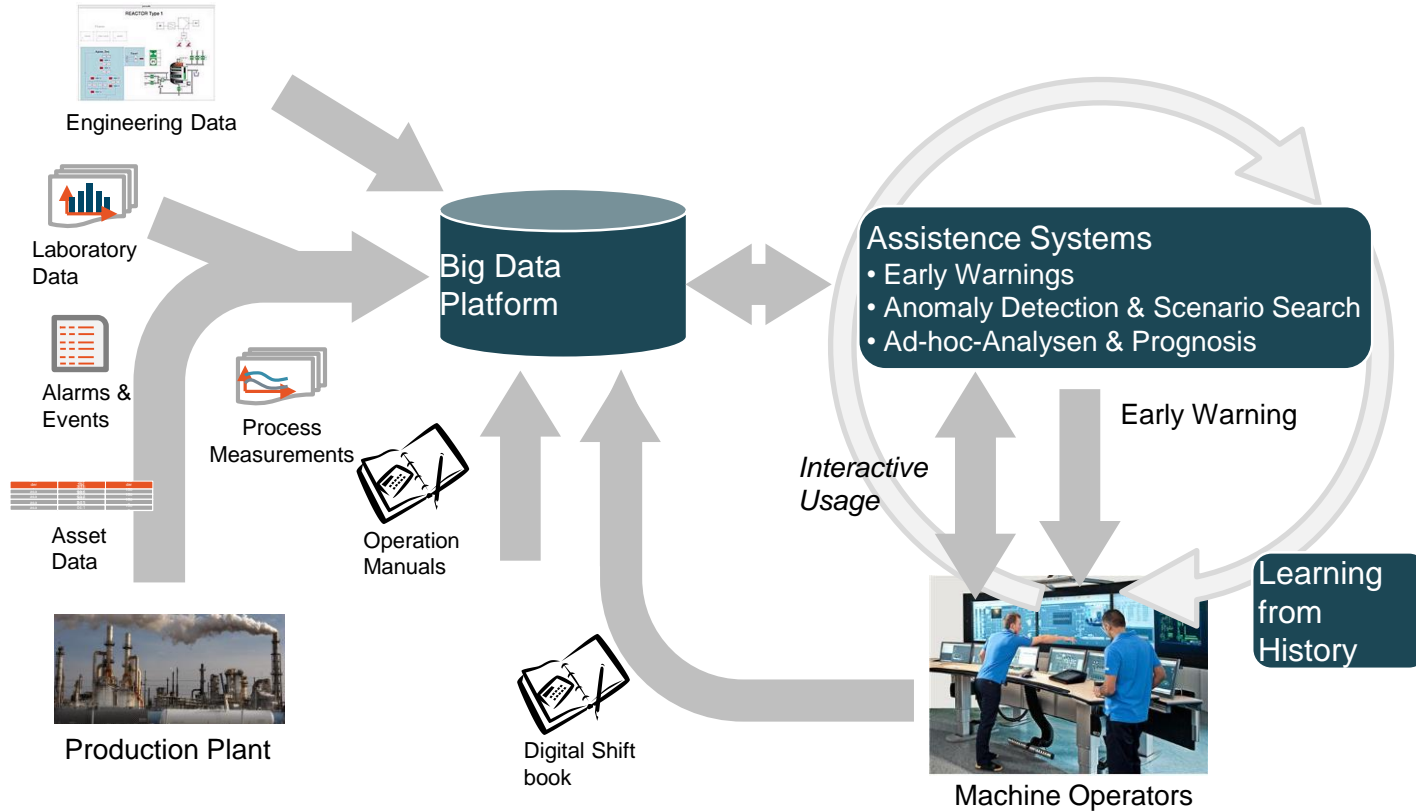


FKZ: 01| S14006A-E

Project Overview

- Design and build a Big Data infrastructure to manage the data of production plants
- Develop methods for early detection and prediction of critical situations
- Develop methods to help users during critical situations or to avoid critical situations
- Build ad-hoc analysis functions to build intervention strategies
- Change from reactive to proactive action
- Use cases:
 - Predicting and prevent critical events
 - Search historic data in support of improved planning

FEE – Data and System Landscape



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



Project “Pro
Mondi – Predicting
Assembly Plans in
Digital Factories”
sponsored by the
German
Government
(BMBF)

SIEMENS



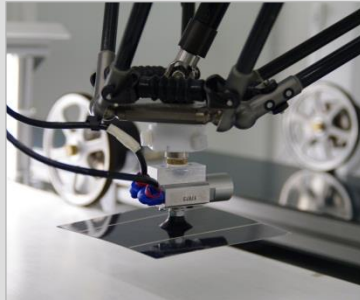
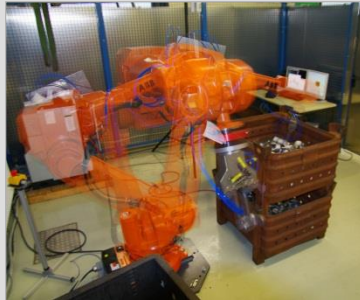
DAIMLER



BETREUT VOM



GEFÖRDERT VOM



Truck Manufacturing – Prediction of Assembly Times & Plans

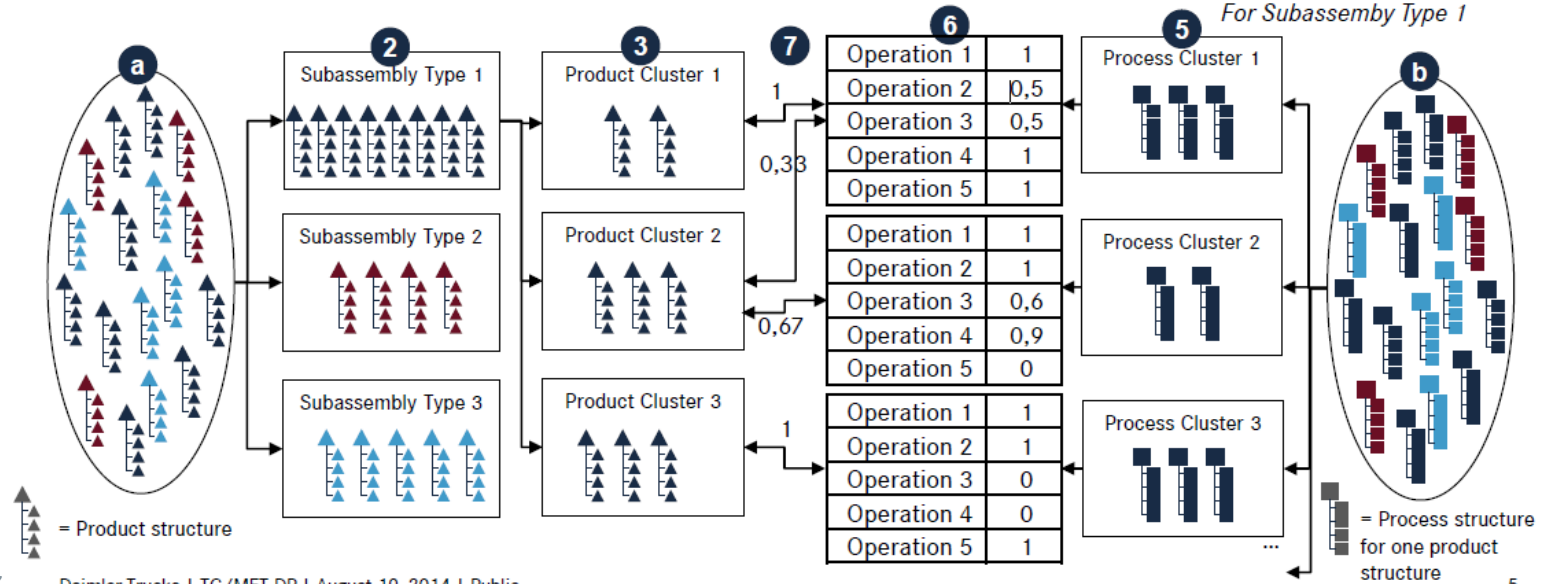


- Uses RapidMiner to predict assembly times and hence assembly costs early in the design process of new truck engines
 - ... so that product designers can faster decide which design options are most cost-efficient, reducing assembly cost as well as design time and costs
- RapidMiner also predicts assembly plans for the new truck engine or engine component designs
 - ... supporting the assembly planer to faster provide final assembly plans for new truck engines and components
- This accelerates the product design and assembly planning phases and to lower the cost, thereby making the whole process faster and less costly
 - As a result, the process becomes cost-efficient even for more and smaller series and less frequent product variations, making the car maker more competitive
- Proof-of-concept demonstrator sponsored by the German government

- **Input Data:**
 - Standard: Linear Feature Vector
 - Here: Complex 3D Structure of 100s or 1000s of components
 - Similarity Metric Based on Sub-Graph Matching Would Be NP-Complete => Not Scalable to Large Data Volumens
- **Output Data:**
 - Standard: atomic value, e.g. class or categorie or regression value
 - Here: complete assembly plan consisting of a long sequence of steps, for which the order is important
- **Solution:**
 - Simplify
 - Decompose to sub-problems

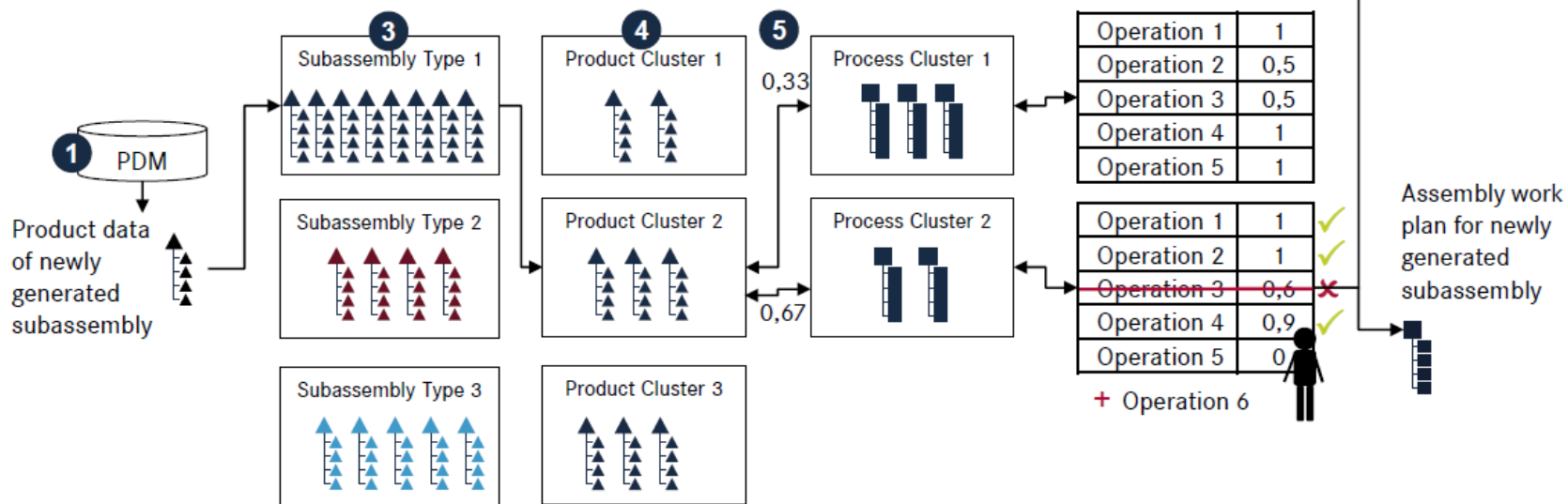
Model training for an automatic proposal of assembly work plans

- a** Product BOMs of different subassembly types
- 1** Product BOM characterization with feature vectors
- 2** Model training for subassembly type classification
- 3** Subassembly type specific product clustering
- b** Bills of processes (BOPs) of different subassembly types
- 4** BOP characterization with PMTS codes and text vectors
- 5** Subassembly type specific process clustering
- 6** Storage of process cluster specific information (incl. process time)
- 7** Generation of a subassembly type specific mapping between product and process clusters



Model application for an automatic proposal of assembly work plans

- 1 Product data extraction out off PDM system
- 2 Data preprocessing (background)
- 3 Classification into subassembly type
- 4 Classification into existing product cluster
- 5 Mapping of product and process cluster (including process cluster specific information)
- 6 Manual adjustment of proposed assembly work plans



Optimizing Product Design, Minimizing Assembly Costs, and Increasing Agility by Predicting Assembly Costs and Plans

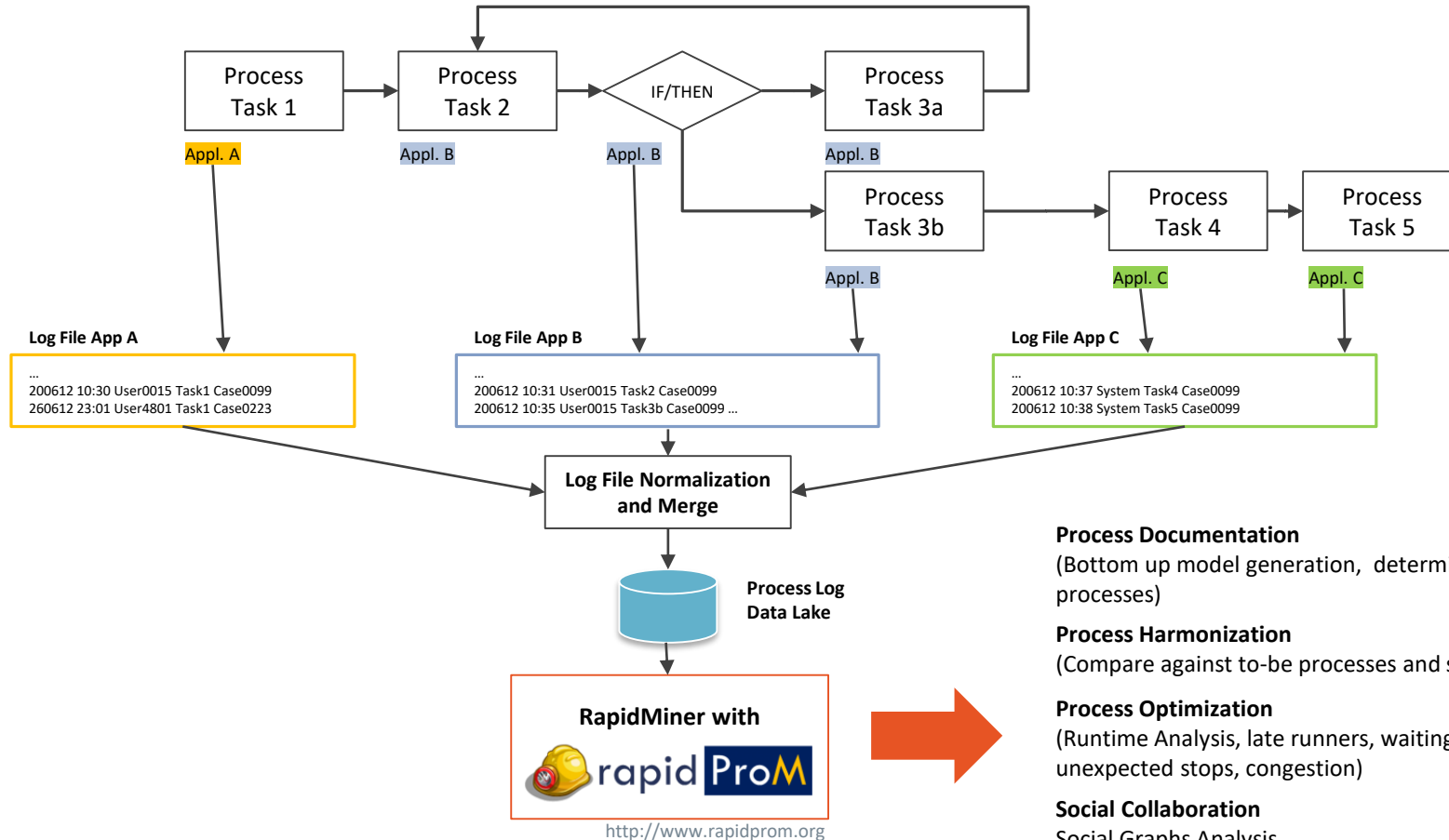
- Competitive Pressure in the Automotive Industry
 - => Faster Product Life Cycles & More Product Variants
 - => Time Pressure & Cost Pressure
- Product Design Decision Have an Impact on Assembly Times & Costs
- Leveraging the Expertise of Experienced Produkt Designers & Assembly Planers Using Machine Learning from Product Designs & Assembly Plans of Previous Products
 - => Support Product Designers & Assembly Planers with Predicted Assembly Times, Costs, and Assembly Plan Recommendations Using Data Mining and Text Mining
 - => Automatically Predict Assembly Plans & Costs Early in the Product Design Process for Alternative Designs to Select Cost-Optimized Design Early in the Process
 - => Saving Time & Resources During Product Design, Assembly Planning, and Assembly
 - => Faster Process, Lower Cost, Higher Agility, Implicit Know-How Transfer
- Use Case at Daimler Trucks for Truck Engines & Components ("Pro Mondi" Project)

Predicting Assembly Plans

- Given the hierarchical design of a new product or variant, predict an appropriate assembly plan and its associated costs.
- Input Data: Product design consisting of components consisting of parts, i.e. multi-level hierarchy of components & parts, highly varying structure, and no one-to-one mapping of parts to assembly steps.
- Need to consider types and numbers of parts and components used, textual descriptions of components and parts, physical measurements, materials, and other properties and features (multi-criteria metrics).
- Desired Output: Assembly plans = sequence of assembly steps.
- => Complex & Challenging Prediction Task
- => Accurate Predictions with RapidMiner
- => Use Case at Daimler Trucks: Assembly Time & Assembly Plan Prediction

- Financial Audits
 - Compliance / regulatory audits
 - Operational audits
 - Transactional services (M&A)
- IT Audits
 - IT Service management
 - Cyber security
 - Systems compliance
 - IT forensic services
- Manufacturing
 - Identifying assembly bottlenecks
- Manufacturing and business processes leave footprints and audit trails
- ... typically log files
- Use process process mining to :
 - Collect
 - Normalize
 - Correlate
 - Analyze
- Unlike BAM this analysis has no defined dimensions
- Use RapidProM from the RapidMiner Marketplace

Process Mining – RapidMiner [with ProM]



gefördert vom:



betreut vom:



„Sozio-technische Gestaltung und Einführung cyber-physischer
Produktionssysteme in nicht F&E-intensiven Unternehmen“

Web-basierte Auswahlhilfe für CPPS-Lösungen + Intelligentes Assistenzsysteme für Angebots-Erstellung und Einkauf

Konsortium:



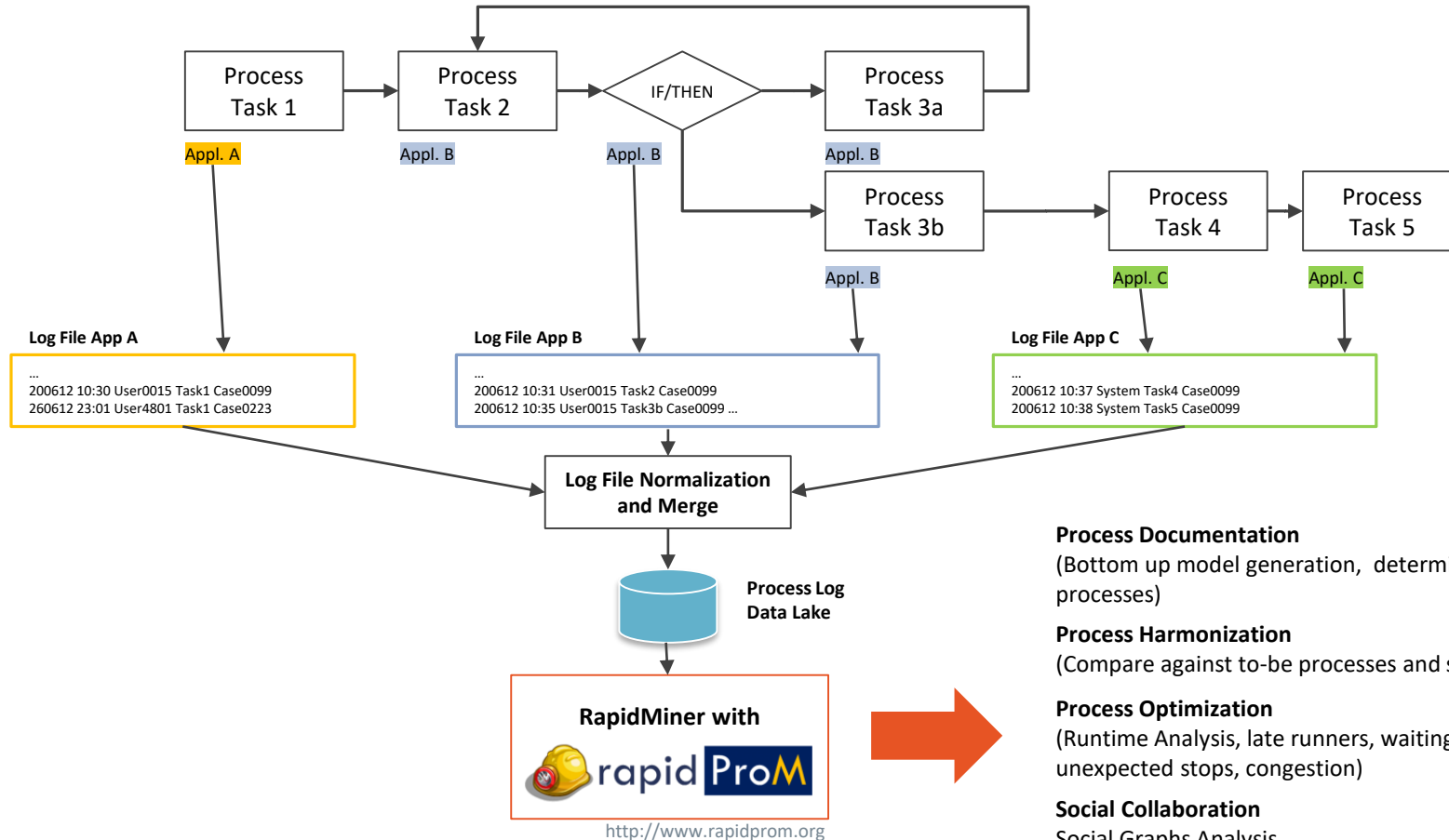
unterstützt

durch:



- Financial Audits
 - Compliance / regulatory audits
 - Operational audits
 - Transactional services (M&A)
- IT Audits
 - IT Service management
 - Cyber security
 - Systems compliance
 - IT forensic services
- Manufacturing
 - Identifying assembly bottlenecks
- Manufacturing and business processes leave footprints and audit trails
- ... typically log files
- Use process process mining to :
 - Collect
 - Normalize
 - Correlate
 - Analyze
- Unlike BAM this analysis has no defined dimensions
- Use RapidProM from the RapidMiner Marketplace

Process Mining – RapidMiner [with ProM]



Text Analytics to Improve Research Effectiveness



The Challenge

- Over 10,000 researchers struggle to communicate, collaborate.
- Research requires a team of "indexers" who read, tag, categorize, and organize 100 years of research.
- As many as 1000 new research papers are submitted each day.

RapidMiner Solution

- Data science team created a text analytics process using RapidMiner, eliminating manual tagging.
- RapidMiner categorizes, tags, and deploys research documents to correct location in the research library.

Outcome

- Reduce millions of cost through improved productivity
- Reduced duplicate efforts, resulting in cost savings



Process Customer Feedback In Multiple Languages To Increase Retention Rates



Challenge: Applying basic voice-of-the-customer-concepts and text analytics to customer feedback in over 60 countries worldwide.

Solution: Use RapidMiner's Platform to detect churn and identify customer service issues regardless of time, location or language.

150,000
customer comments and tweets in almost every language processed on RapidMiner

Data Science Hero Spotlight

"Business executives, who hold the power to allocate text analytics resources, are beginning to see and realize the benefits to help better focus and solve business problems."

-- Han-Sheong Lai

Director of Operational Excellence & Customer Advocacy



Accelerate

Process massive amounts of text at high speed



Connect

Analyze multiple silos of global customer data



Simplify

Automatically determine intent-to-churn



Quickly Prototype Analytics Models for Under Armour Challenge Wearables Data



Challenge: Quickly prototype analytics processes for Under Armour wearable data, for the Under Armour39 Challenge.

Solution: Use RapidMiner's code free, drag and drop GUI to quickly design 11 analytics processes, iterate them for optimization, and win the challenge.

1.8M
data points analyzed,
per hour, by the Under
Armour39 wearable

Data Science Hero Spotlight

"RapidMiner is extremely powerful, has the best operators, and can handle Big Data from wearables. It also allows us to rapidly prototype sophisticated analytics, machine learning and classification applications, saving time and money."

-- Kevin Logan
CEO



Accelerate

Prototype multiple analytics processes quickly and easily



Connect

Analyze Big Data from wearables devices



Simplify

Use code free, drag and drop GUI for analytics



Search Millions of Patents Online and Automatically Mine Image Data

Challenge: Search millions of patents online and automatically mine image data for applicable information.

Solution: Use RapidMiner text and image mining to quickly and easily identify several thousand images of interest.

1M+
detailed patent records
mined online,
including images

Data Science Hero Spotlight

“Some years ago (the patent team) had tried a dedicated patent classification tool that didn’t work - RapidMiner does. It provides a framework for substantially reducing the time it takes us to find interesting patents.”

-- Thomas Hartmann
Business Engineer



Accelerate

Automatically mine millions of online patent images



Connect

Search through a wide variety online data sources



Simplify

No programming required to connect insights to action

Drive Broadcast Revenues and Customer Retention with Streaming, Real-Time Analytics

Challenge: Better understand TV viewing habits to prevent churn and optimize advertising.

Solution: Process streaming Big Data from three million TV viewers, in real-time, to make program content recommendations and target advertising.

<5s

time to generate high value activities based on predictive analytics

Data Science Hero Spotlight:
“RapidMiner allows us to leverage Big Data, in real-time, for the TV industry.”

-- **Avi Bernstein**
Professor at the University of Zurich, Department of Informatics



Accelerate

Personalized
recommendations < 5s



Connect

Stream and analyze from set-top
boxes, mobile devices and PCs



Simplify

Code free design of
streaming analytics



New Major Research Projects 2019 - 2021: Data Stream Processing, Industry 4.0, IoT, Web Mining

AKKORD: Smart Analytics Services Platform for Industry 4.0 & IoT Applications

- Partners include Miele, Volkswagen, IPS TU Dortmund University, etc. (Industry 4.0 Cloud).

DaPro: Optimization of Industrial Production Processes with Data Mining (here with example processes from the beer brewing industry)

- Partners include beer brewing companies Bitburger Group and Augustiner, machine builder Syskron/Krones, and IPS, TU Dortmund University.

INFORE: Real-Time Large Scale Data Stream Processing, Monitoring, and Analytics

- Partners include NATO, Maritime Traffic Monitoring, Medical Partners, etc.

TechRad: Automated Web Mining & Monitoring for a Technology Trend Radar

- Partners include eGo Car (Electric Car Manufacturer), RWTH Aachen University, etc.



DaPro

Data Driven Process Optimization in the Brewing Industry

Started 01/2019





INFORE

Interactive Extreme-Scale Analytics and Forecasting

Started 01/2019



INFORE

Interactive Extreme-Scale
Analytics and Forecasting

- ATHENA - Research and Innovation Center (coordinator)
- NCSR Demokritos
- Barcelona Supercomputing Center
- The Centre for Genomic Regulation
- Spring Techno
- NATO Centre of Maritime Research and Experimentation (CMRE)
- MarineTraffic



- **Goal:**

Handling massive data flows from multiple sources and allow real-time interactive analytics

- **Use cases:**

Data Streaming, High Performance Computing, *In-Silico* simulation

- **Verticals:**

Finance, Maritime (ship movement), Medical (cancer research)

High-Level Overview

- Funding: European Union, H2020
- Duration: 3 years (2019-01 – 2021-12)
- Topics & Goals:
 - Analyze huge datasets and pave the way for real-time, interactive extreme-scale analytics and forecasting
 - RapidMiner is the main driver for the technical development
 - Three use cases:
 - Monitoring global marine traffic
 - In-Silico analysis of tumor growth
 - Real-Time analysis of online trading

- What do you customers like/dislike about my products & services and about those of my competitors?
(e.g. Android SmartPhone X vs. Apple iPhone Y)
- Web Crawler: Automatically collect opinion from the web: social media like Facebook and Twitter, internet discussion groups, product reviews (e.g. Amazon), etc.
- Automated Text Classification: Automatically focus on relevant pages and statements based on content and sentiment
- Determine top words and phrases indicating positive / negative sentiment, i.e. reasons for like / dislike, e.g. preferred product features or service issues or pricing policies, etc.
=> Better understand customers & market
=> Tailor product better to demand

- **Task:**
 - How many cars of each model and configuration will be sold in a given future time period?
- **Challenges:**
 - Large number of products and configuration variants
- **Solution:**
 - Consider clusters of products and/or configurations
 - Consider customer information requests and online configurations
 - Consider overall market trends (economic up-/down-turns etc.)
 - Consider social media discussions and e-mail communications to detect trends and preferences
 - Analyse historic sales data for trends and patterns (seasonal, colors, etc.)

Demand Forecasting & Supply Chain Mgt.

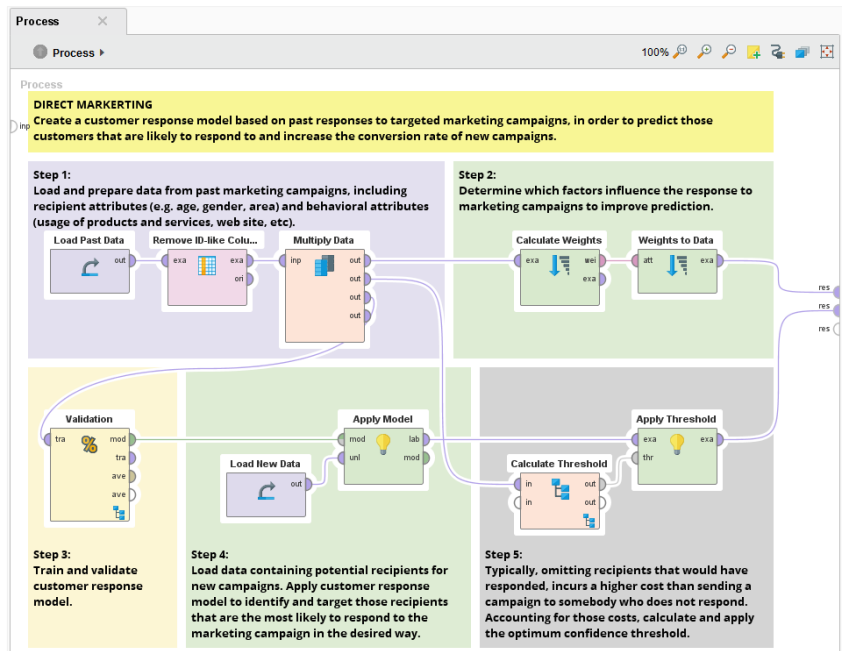
- How often will each product / service be sold?
- Customer: Leading Telecom Equipment Supplier
- Solution: Demand Forecasting
 - => Optimize Planning & Time of Orders (Prices)
 - => Optimize Supply Chain Management
- Related: Price Forecasting & Dynamic Pricing
- Related: Network Load Forecasting, Bottleneck Prediction

Product Market Fit & Regional Issues

- **Task:**
 - Analyse Product Failures & Issues
 - Detect Regional Issues & Opportunities
- **Challenges:**
 - Long-Term Task & Data Collection across Supply Chain & Product Lifecycle
 - Data Silos
- **Solution:**
 - Comprehensive Data Collection
 - Breaking Data Silos

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

RapidMiner **Server** & **Real-Time Scoring**



Operationalization & Collaboration Management

Team Collaboration

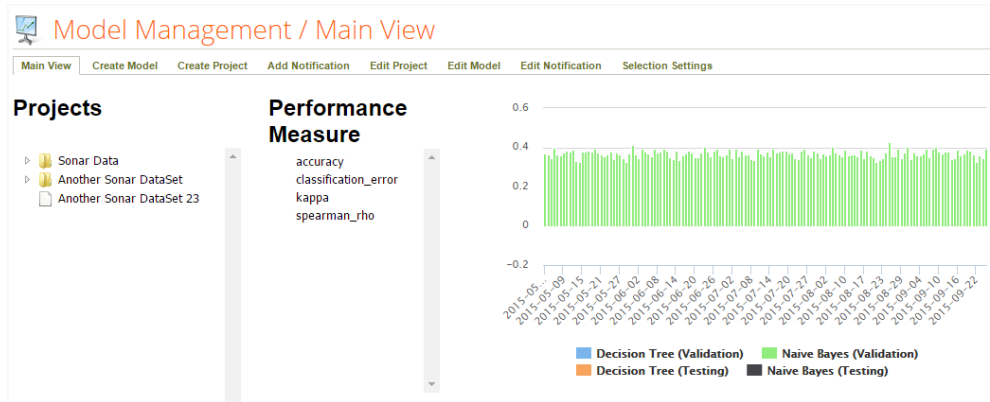
Central repository
facilitates sharing of data
sources, analytic
processes & best
practices

Frictionless Operationalization

Flexible execution options
streamline deployment,
maintenance & embedding of
analysis

Dynamic & Continuous Model Management

Individual and customizable
processes to check for accuracy
drifts or shifts



#1 Data Science Platform & Marketplace

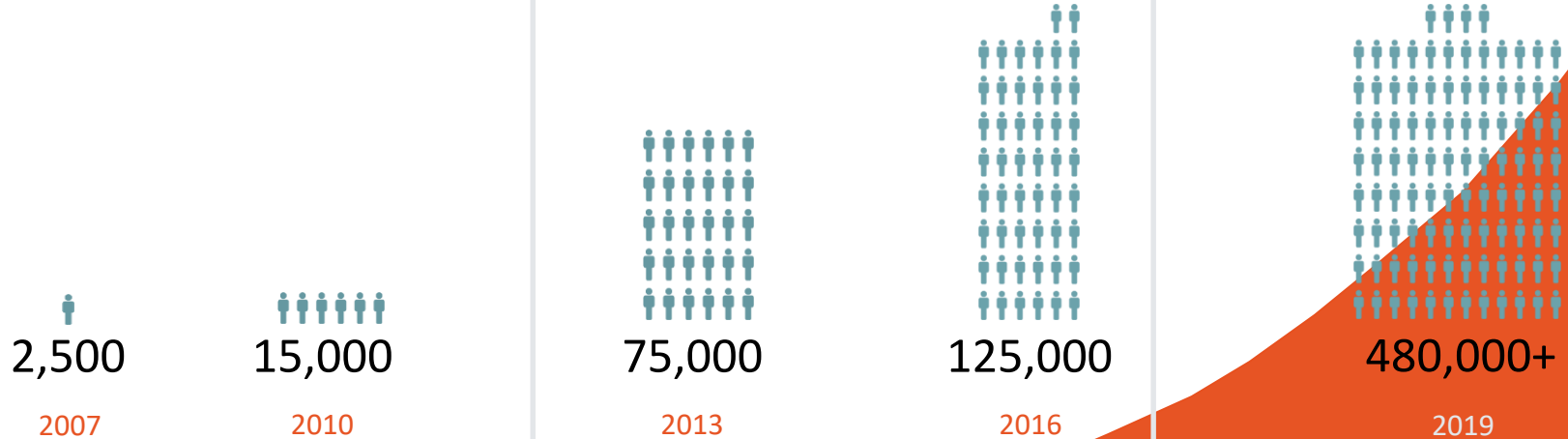
#1 Open Source Community



EMERGING

EXPANDING

INTENSIFYING



Number of Registered Users

<http://community.rapidminer.com/>

Signature Customers



PayPal[™]



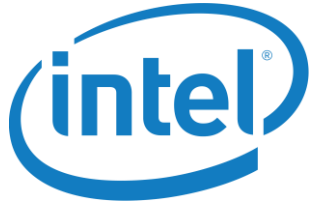
AMERITRADE



CALIBER
HOME LOANS



 **BARCLAYS**



DAIMLER



SANOFI



DIRECTV.



■ ■ ■ **T** Deutsche
Telekom

P&G



Selected RapidMiner Customers & Partners

- Automotive: BMW, Daimler, Jaguar Land Rover, Nissan, Volkswagen
- Automotive Components: Bosch, Conti (Continental), ThyssenKrupp Presta
- Airplane Manufacturers & Aviation: Airbus, Embraer, Lufthansa
- Metal & Steel Industry: Arcelor Mittal, Bosch, ThyssenKrupp Steel, Mannesmann Salzgitte, Deutsche Edelstahlwerk (DEW)
- Electronics & Manufacturing Related: ABB, General Electrics, Intel, KHS, Miele, MTM, Schneider Electric, Siemens
- Chemical Industry: AkzoNobel, BASF, Covestro, Dow, Ineos, PCK, Sanofi
- Energy & Oil Industry: CEPRI, Chevron, E.ON, RWE, Saudi Aramco, TOTAL
- Telecom Industry: Deutsche Telekom, mobilkom austria (A1.net), Nokia, SmartSoft (OEM), SwissCom, Telenor, T-Mobile International
- Customer Experience Management & Churn Prevention: PayPal, T-Mobile International, Telenor, mobilkom austria (A1.net), Nokia, SwissCom, Banks, Financial Service Companies, Insurance Companies



THANKS.



Ralf Klinkenberg

rklinkenberg@rapidminer.com

www.RapidMiner.com