

March 2019

# **Automated extraction of balance sheet information**

KNIME Summit 2019

- 1 Smart Data at LBBW
- 2 Automated extraction of balance sheet information

# LBBW: A brief overview



## Medium-sized universal bank

LBBW is a medium-sized universal bank with deep roots in Baden-Württemberg



## Strong partner for our customers

We are a partner for SMEs, listed corporations, institutional customers, savings banks and private customers



## Data and facts



Head offices: Stuttgart, Mainz, Karlsruhe and Mannheim



Group employees December 2017: Approximately 10,300



Total assets December 2017: EUR 238bn



## Specialized Group subsidiaries

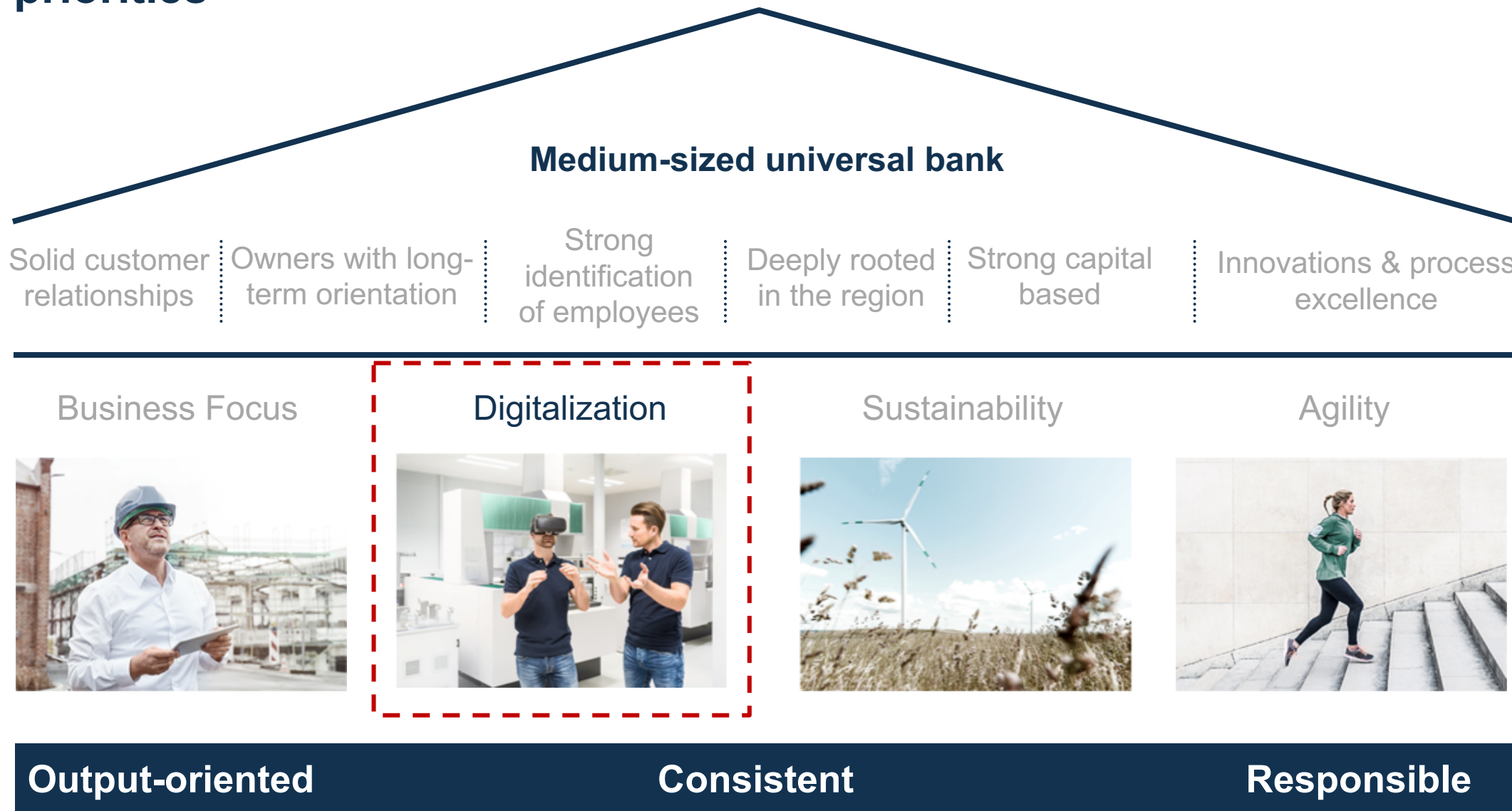
**LB≡BW Asset Management**   **LB≡BW Immobilien**

**MKB**  **Bank**

 SüdLeasing

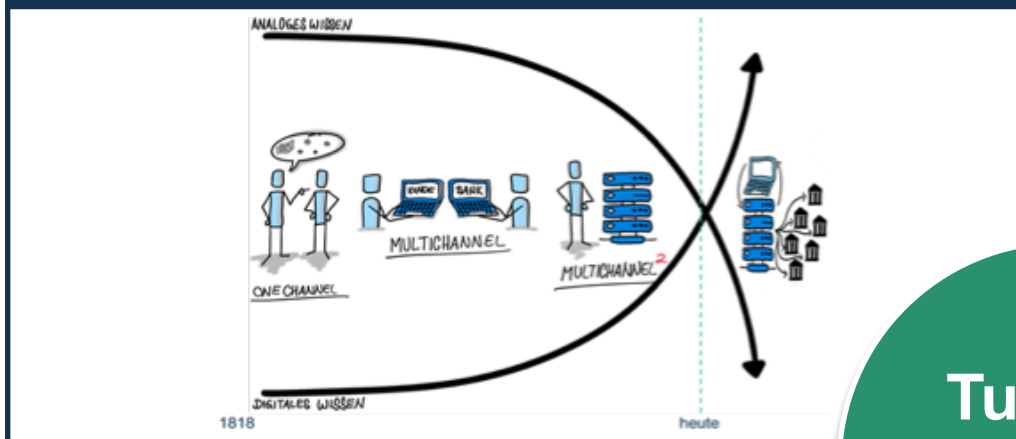
SüdFactoring 

# Our four priorities



# The Smart-Data Concept

## Digital touch points



## Archived states



Turning  
Data into  
Value







Mathematics and statistical methods



Better Decisions

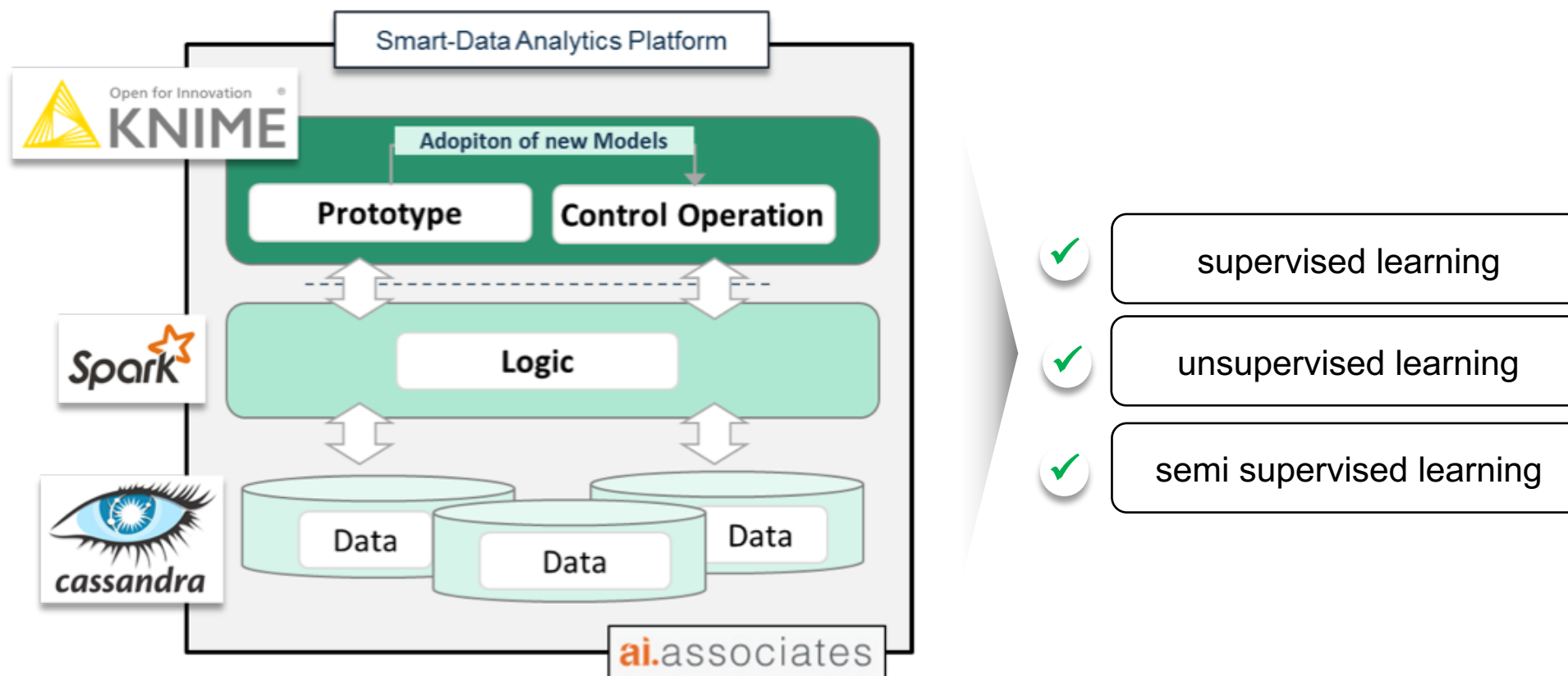
# Achieve impact at all touch points

People	Markets	Companies	Cooperations
			
<ul style="list-style-type: none"> <li>▪ Product affinity</li> <li>▪ Networks and Relations</li> <li>▪ Forecast behaviour purchasing</li> <li>▪ Fraud detection and money laundering</li> <li>▪ Process automation</li> <li>▪ Risk prediction</li> <li>▪ ...</li> </ul>	<ul style="list-style-type: none"> <li>▪ Risk prediction</li> <li>▪ Sensitivity analysis</li> <li>▪ Market and competition analysis</li> <li>▪ Network analysis and behavior forecasting</li> <li>▪ Trend analysis</li> <li>▪ ...</li> </ul>	<ul style="list-style-type: none"> <li>▪ Balance sheet analysis</li> <li>▪ Implementation Industry 4.0</li> <li>▪ Machine-Data</li> <li>▪ Document management</li> <li>▪ Fraud detection and money laundering</li> <li>▪ Transaction Management for Data</li> <li>▪ ...</li> </ul>	<ul style="list-style-type: none"> <li>▪ „Data Analytics as a Service“ for our corporate customer</li> <li>▪ Scientific cooperation with Universities and the KIT<sup>1</sup></li> <li>▪ Saving Bank Finance Group</li> <li>▪ Technical Cooperation with the HLRS<sup>2</sup></li> <li>▪ ...</li> </ul>

<sup>1</sup>Karlsruher Institut für Technologie

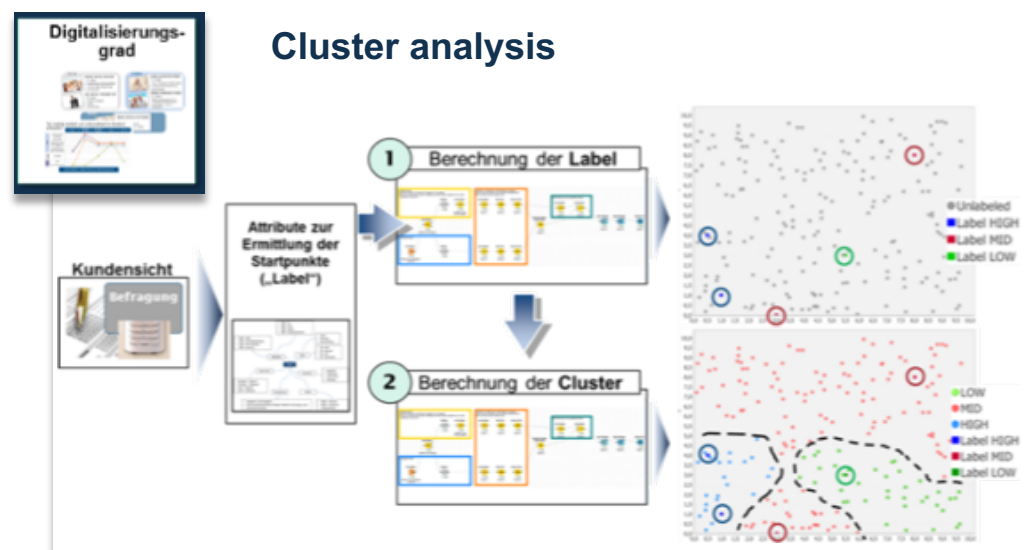
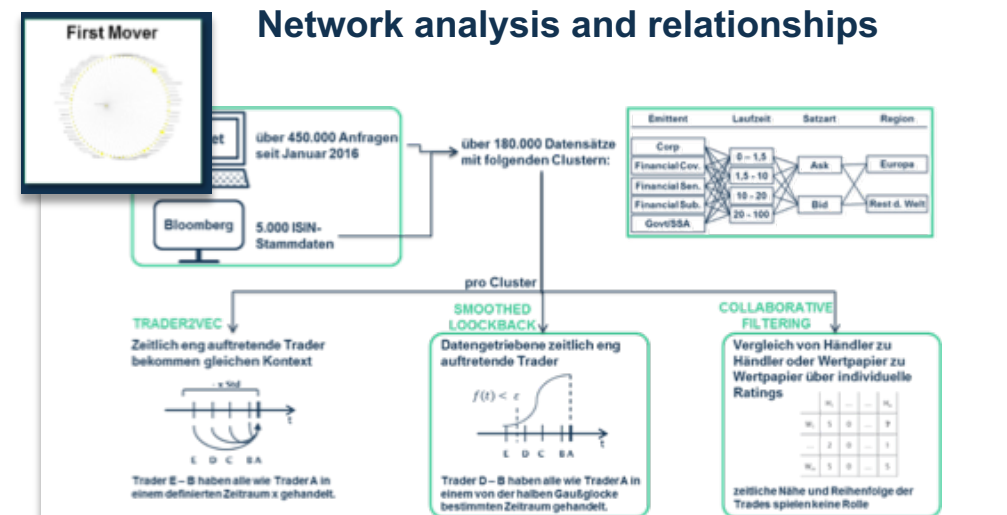
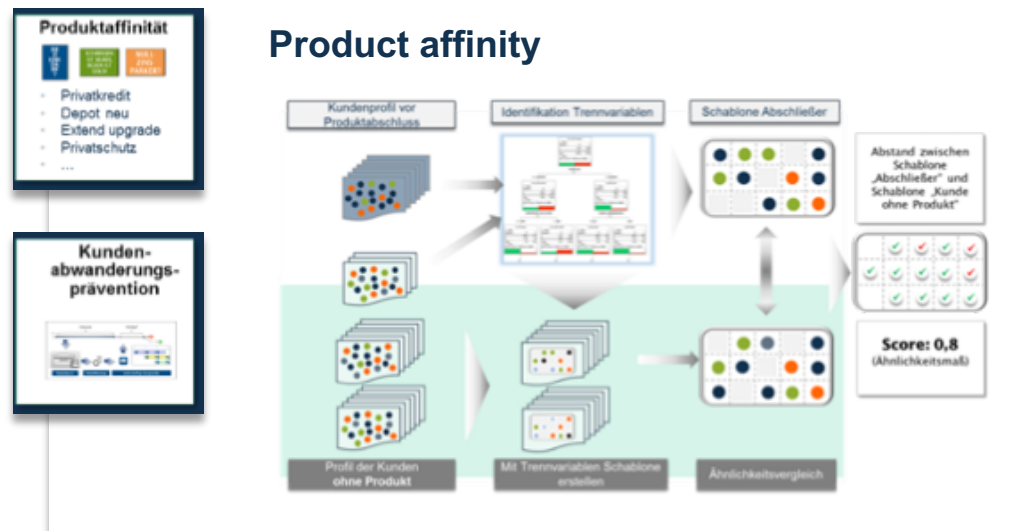
<sup>2</sup>Höchstleistungsrechenzentrum Stuttgart

# One technical platform for all business units



**Mission: implement Artificial Intelligence in the business processes**

# Some examples





- 1 Smart Data at LBBW
- 2 Automated extraction of balance sheet information
  - data
  - goal
  - pilot project in 4 use cases
    - method
    - results
  - next steps

# Goals of Balance Analysis



## Motivation

- departments like Research and Advisory need all these information to provide their clients with tailor-made and targeted advice
- important for investments and comparison with other companies in the same peer group

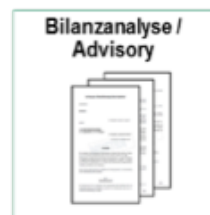
## Previous manual approach

- reading thousands of balances manually
- needing all information on each page
- extract special key figures, hints and facts about risks, opportunities and other KPIs of a balance and relatable companies in the same branch

## Benefits of doing it automatically

- faster
- cost saving
- concentrated on what departments do
- no need to read everything

# Pilot Project – In 4 Use Cases



2 Management Summary

2.1 Ziel der Publikation

2.2 Erkenntnis aus dem Big Data-Lernprozess

2.3 Erkenntnis aus dem Big Data-Lernprozess

2.4 Erkenntnis aus dem Big Data-Lernprozess

Classify branch/industry

Extract key figures from tables (within pdf)

Extract key figures from plain text

Extract text sections containing company risks or opportunities

- **Why** extract key figures?
- **Examples:** revenue, EBIT (earnings before interest) etc.
- **How?**

Models learn **word patterns** in order to identify and distinguish key figures, text section candidates and industries.

# Extract key figures from balance tables

## Benefits & Drawbacks

- Almost all key figures we want
- Nicely structured information ...
- ...but difficult to extract from PDF files.
- Almost impossible to identify after parsing PDF files into plain text.
- Common KNIME Parser Nodes not applicable.

### Erfolgszahlen:

	01.01.2017 - 31.12.2017 Mio. EUR	01.01.2016 - 31.12.2016 <sup>1</sup> Mio. EUR	Veränderung	
			Mio. EUR	in %
Zinsergebnis	1 587	1 669	-82	-5
Risikovorsorge im Kreditgeschäft	-92	-51	-41	81
Provisionsergebnis	534	527	7	1
Ergebnis aus erfolgswirksam zum Fair Value bewerteten Finanzinstrumenten	219	146	73	50
Finanzanlage- und at-Equity-Ergebnis	163	195	-32	-17
Sonstiges betriebliches Ergebnis	101	101	0	0
<b>Nettoergebnis (nach Risikovorsorge)</b>	<b>2 511</b>	<b>2 586</b>	<b>-75</b>	<b>-3</b>
Verwaltungsaufwendungen	-1 824	-1 814	-11	1
Garantieprovision Land Baden-Württemberg	-61	-93	32	-34
Aufwendungen für Bankenabgabe und Einlagensicherung	-69	-71	2	-2
Wertminderung Goodwill	0	-379	379	-100
Restrukturierungsergebnis	-41	-87	47	-54
<b>Konzernergebnis vor Steuern</b>	<b>515</b>	<b>142</b>	<b>373</b>	<b>&gt; 100</b>
Ertragsteuern	-97	-131	34	-26
<b>Konzernergebnis</b>	<b>419</b>	<b>11</b>	<b>408</b>	<b>&gt; 100</b>

Aus rechnerischen Gründen können in dieser und den nachfolgenden Tabellen Rundungsdifferenzen auftreten.  
<sup>1</sup> Anpassungen Vorjahreswerte (siehe Note 2).

## Solution: New PDF – Parser Node

**AI.Associates**  
PDF Parser



Read Input  
PDF

# Features: Words in context

- Words close by that frequently cooccur
- Create word vectors

Nettoergebnis	EBIT	Ertragssteuern	Wertminderung	Umsatz

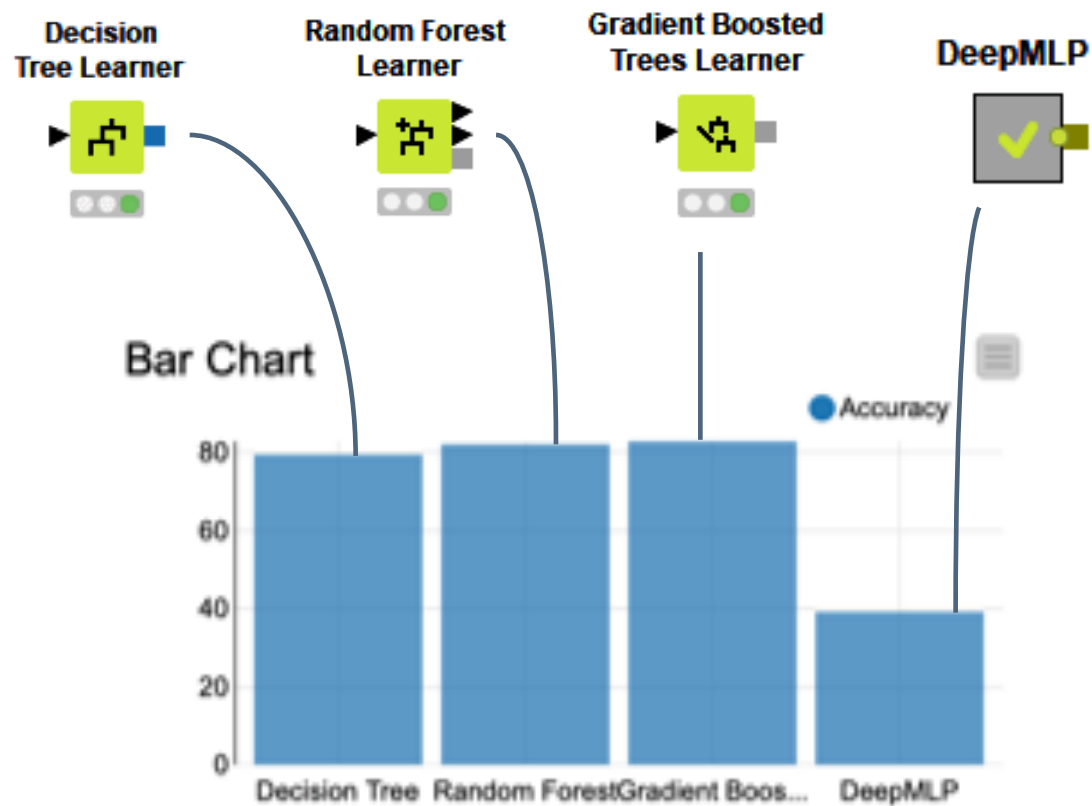
**Erfolgszahlen:**

	01.01.2017 - 31.12.2017 Mio. EUR	01.01.2016 - 31.12.2016 <sup>1</sup> Mio. EUR	Veränderung	
			Mio. EUR	in %
Zinsergebnis	1 587	1 669	-82	-5
Risikovorsorge im Kreditgeschäft	-92	-51	-41	81
Provisionsergebnis	534	527	7	1
Ergebnis aus erfolgswirksam zum Fair Value bewerteten Finanzinstrumenten	219	146	73	50
Finanzanlage- und at-Equity-Ergebnis	163	195	-32	-17
Sonstiges betriebliches Ergebnis	101	101	0	0
<b>Nettoergebnis (nach Risikovorsorge)</b>	<b>2 511</b>	<b>2 586</b>	<b>-75</b>	<b>-3</b>
Verwaltungsaufwendungen	-1824	-1814	-11	1
Garantieprovision Land Baden-Württemberg	-61	-93	32	-34
Aufwendungen für Bankenabgabe und Einlagensicherung	-69	-71	2	-2
Wertminderung Goodwill	0	-379	379	-100
Restrukturierungsergebnis	-41	-87	47	-54
<b>Konzernergebnis vor Steuern</b>	<b>515</b>	<b>142</b>	<b>373</b>	<b>&gt; 100</b>
Ertragsteuern	-97	-131	34	-26
<b>Konzernergebnis</b>	<b>419</b>	<b>11</b>	<b>408</b>	<b>&gt; 100</b>

Aus rechnerischen Gründen können in dieser und den nachfolgenden Tabellen Rundungsdifferenzen auftreten.  
1 Anpassungen Vorjahreswerte (siehe Note 2).

# Model training & Evaluation

- Train & evaluate best performing model
- Deep Learning is not the solution for everything



Show 10 entries Search:

<input type="checkbox"/>	Company	Page Index	Category	Quantity
<input type="checkbox"/>		70	Abschreibungen	-8193
<input type="checkbox"/>		98	CAPEX	47
<input type="checkbox"/>		70	Ergebnis vor Steuern	16631
<input type="checkbox"/>		71	Kasse / Flüssige Mittel	43144
<input type="checkbox"/>		71	Kurzfristige Finanzverbindlichkeiten	5500
<input type="checkbox"/>		71	Langfristige Finanzverbindlichkeiten	5500
<input type="checkbox"/>		71	Operativer Cashflow	20265
<input type="checkbox"/>		70	Umsatz	321590

Showing 1 to 8 of 8 entries Previous 1 Next

## Extract key figures from free text

### Why free text search?

- Not necessarily all key figures present in tables
- Additional contextual information
- Why not?

ch der tatsächlichen Entwicklung bestehenden und teils neuen Risiken für ihre Projektion unterstellt die Regierung explizit, dass es im Prozess zu keinem sprunghaften Anstieg kommt, beispielsweise im Zuge der Entwicklungen. Darüber hinaus wird davon ausgegangen, dass der Finanzsektor stabil bleibt und keine weitere Eskalation geopolitischer Risiken aufgrund der bisherigen Maßnahmen zur Stabilisierung des Finanzsystems zu erwarten ist. Der Finanzsektor wird jedoch nicht damit gerechnet, dass die bisherige positive Wirkung in den für adesso bedeutsamen Jahren 2019 massiv verschlechtert

mente sowie den internen Überlegungen erwartet das Management für das Geschäftsjahr 2018:

> Für die Umsatzerlöse wird mit einer Steigerung um mindestens 9 % auf mindestens 350 Mio. € gerechnet.

> Für das operative Ergebnis, EBITDA, wird mit einer Steigerung um mindestens 9 % auf mindestens 27,7 Mio. € gerechnet.

> Für die operative EBITDA-Marge wird wie im Vorjahr ein Mindestwert von 7,5 % erwartet. Die EBITDA-Marge soll nach Möglichkeit jedoch weiter in Richtung des langfristigen Zielkorridors gesteigert werden.



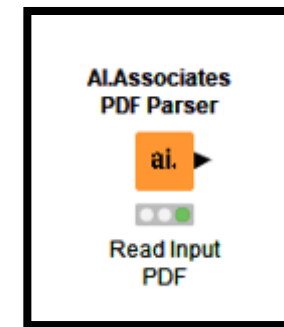


# Identify structure of balance

- Infer reading directions from word positions



VS.



Thanks for making it possible



# Model evaluation & Overall accuracy

Relevance Prediction



Precision (%)	Recall (%)
80,6	73,5

Precision (%)	Recall (%)	Accuracy (%)
58,2	65,4	≈ 60

4 model classes

Predict Key Figure X



Predict Key Figure Y

Ø Precision (%)	Ø Recall (%)
76,2	78,5

Predict Key Figure Z



## Ways to improve

- More training data for under represented classes (only 100 training sentences available)
- Improve iteratively (e.g. with Active Learning)

Demo



Open for Innovation <sup>®</sup>  
**KNIME**  
WebPortal  
Version 4.7.3

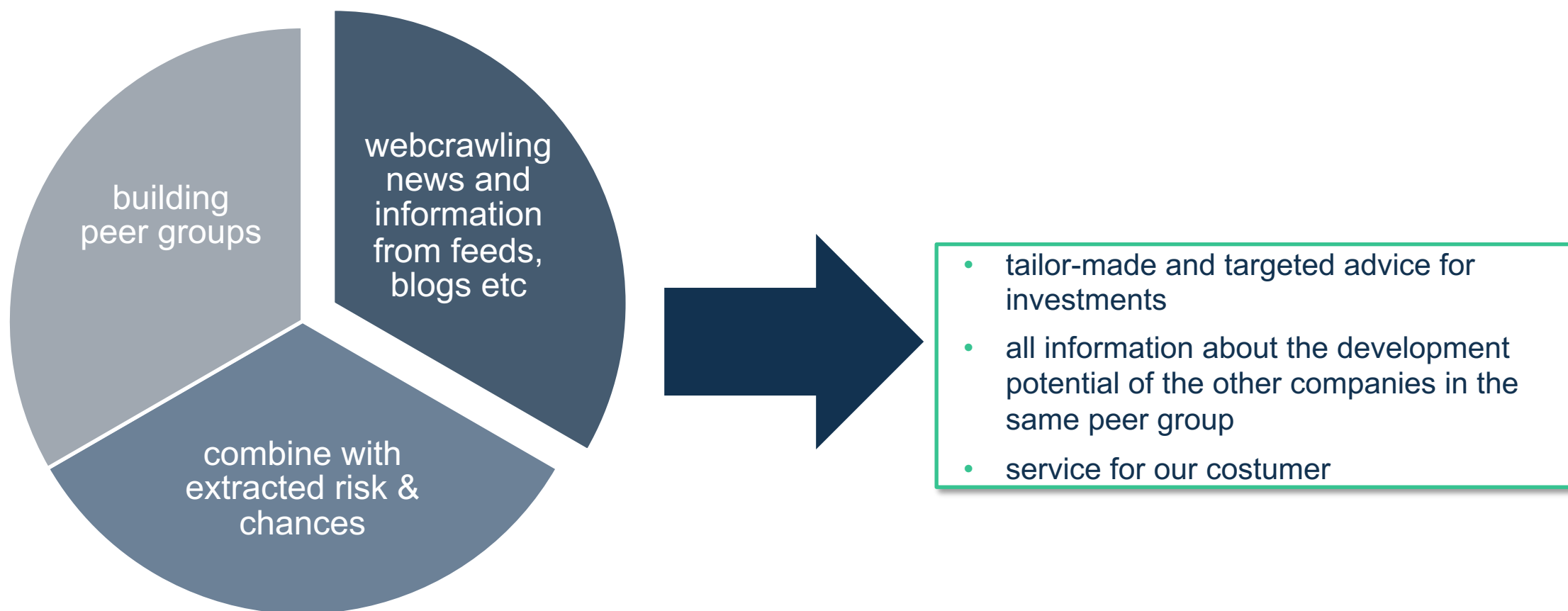
Username

Password

Login



## Next steps



# Thank you for your attention – any questions?



**Michael Höhn**  
Ai Associates

[mh@ai.associates](mailto:mh@ai.associates)



**Christine Diegel**  
2911 Data Analytics

[christine.diegel@LBBW.de](mailto:christine.diegel@LBBW.de)



**Kilian Retter**  
2911 Data Analytics

[kilian.retter@LBBW.de](mailto:kilian.retter@LBBW.de)