AI for product development
- working examples -
The Cornelsen Group at a glance

1,500 employees
24,000 products and services
20,000 teachers
80% of German teachers working with Cornelsen
8 locations
> 1,000 further education courses and workshops/p. a.
for all types of schools and all subjects

many successful brands:
VP Data & Analytics

- Big data & machine learning enthusiast and technologist
- 12+ years experience in IT & AI management:
  - Jamba, mobile startup
  - PAREXEL, pharma
  - readybank, online bank
  - Elsevier, publisher science
  - Cornelsen, publisher education
- Hands-on projects with machine learning (Python / R / Java / H2O / Tensorflow), big data mining (Spark / HPCC / HDFS), predictive analytics (SAS), business intelligence (Power BI / Qlik / D3 / SQL databases)

E-Mail: paul.hellwig@cornelsen.de  Linkedin: https://www.linkedin.com/in/paulhellwig/  Xing: https://www.xing.com/profile/Paul_Hellwig
We started several beacon projects for working with data & AI

**Towards new products**

- Online assessments
- **AI-tools to support adaptive learning**
- Speech recognition
- **Ontologies & Knowledge graphs**
- Handwriting recognition
- **NLP & topic modeling**

**Improving products**

- Personalized website content & search
- **Continuous product optimization with usage analysis**

**Saving money**

- Image usage optimization
- Customer journey analytics
- Marketing & sales targeting
- Business intelligence, process KPIs & dashboards
beacon project example 1:
continuous product optimization
Our question

In which sequence do learners use the contents?

→ We use *process mining* to build a fuzzy model of a graph.

*Nodes*: content blocks

*Edges*: time sequence
Usage patterns

**Group A**
Reads *theory* and does *exercises*. Skips *examples*.

**Group B**
Does *exercises* only

**Gruppe C**
Does *exercises*, but interrupts for reading *theory*

**Group ZZ**
Totally jumps around

**Contents in mathematics**
In basic modules: *Exercises are used more than theory or examples*
In complex modules: *Trend changes, theory increases in popularity*
In-depth examples

**Exercise formats with low cancellation**

- One answer correct
- Short description
- Concise feedback

**Exercise format with high cancellation**

- Multiple answers correct
- Long description
- Long feedback
beacon project example 2: NLP & topic modeling
Identifying text types to tailor style checks & corrections

Batch of ~100k texts from 24hrs

A) NLP pre-processing
   - tokenize, remove stop words, stemming
   - calculate bag of word representation (TF-IDF)
   ...

B) Topic modeling
   - Latent Dirichlet Allocation (LDA)

C) Cluster texts
   - k-means

Checks tailored to text type
B) Examples for topics modeled from texts

- "arbeit", "bereich", "unternehmen", "erfahr", "person", "mitarbeit", "tätig", "aufgab", "projekt", "bewerb" ➔ Topic: "jobs"


C) Text clusters (k=10)

- Cluster-ID 1: "Other" (9154 documents)
- Cluster-ID 2: "job applications" (3908 documents)
- Cluster-ID 3: "job applications" (6068 documents)
- Cluster-ID 4: "job applications" (4387 documents)
- Cluster-ID 5: "job applications" (3046 documents)
- Cluster-ID 6: "job applications" (4793 documents)
- Cluster-ID 7: "patients" (2427 documents)
- Cluster-ID 8: "patients" (1814 documents)
- Cluster-ID 9: "patients" (1820 documents)
- Cluster-ID 10: "patients" (1970 documents)
Word cloud for cluster „job application“