

Extraction of activity fields from German online job posting – Model Card and Dataset Card

Method Report

Project: [Teilqualifikationen - Mehr Fachkräfte in jeder Region](#) (Bertelsmann Stiftung)

Version

1.0.0

Author

Johannes Müller (johannes.mueller@and-effect.com)

Erstellt am

16 March 2023

Letzte Änderung am

16 March 2023

More than 80 % of companies are willing to hire people without vocational qualifications if necessary. This is mainly due to the high degree of specialization in many companies. Especially in times of crisis like the Corona pandemic, people without vocational qualifications are affected by the changes in the labour market. They become unemployed more quickly and have a harder time finding new employment after the crisis. To make the skills of these ‘skilled workers without a vocational qualification certificate’ visible, verifiable activity fields or part qualifications help. Activity fields (German: “Teilqualifikationen”) are subsets of skills or competencies which together form the full qualifications gained by a two- or three-year vocational training (“Berufsausbildung”).

The activity field extraction pipeline can extract mentions of these concepts in German online job ads. The extraction pipeline is built for 27 occupation groups and can be used for 83 standardized occupations in the German Klassifikation der Berufe 2010.

For each occupation group an individual model was trained. Each model consists of two components: A rule-based span extractor and a statistical span extractor. Both components suggest spans in the document that relate to one or more activity fields.

The pipeline is one unified model that combines all 27 individual models. For each document to be classified the ID of the occupation has to be provided in order for the model to decide which of the 56 components to activate for prediction. The suggestions from all active components are merged for the final predictions. Subsequently, all predictions are aggregated on the document level to give a multi-class text classification result.

[The Dataset Card](#) can be found below

Table of Contents

Model Card

- [Model Details](#)
 - [Model Description](#)
- [Uses](#)
 - [Direct Use](#)
 - [Downstream Use \[Optional\]](#)
 - [Out-of-Scope Use](#)
- [Bias, Risks, and Limitations](#)
 - [Recommendations](#)
- [Training Details](#)
 - [Training Data](#)
 - [Training Procedure](#)
 - [Preprocessing](#)
 - [Speeds, Sizes, Times](#)
- [Evaluation](#)
 - [Testing Data, Factors & Metrics](#)
 - [Testing Data](#)
 - [Factors](#)
 - [Metrics](#)
 - [Results](#)
- [Model Examination](#)
- [Environmental Impact](#)
- [Technical Specifications \[optional\]](#)
 - [Model Architecture and Objective](#)
 - [Compute Infrastructure](#)
 - [Hardware](#)
 - [Software](#)
- [Model Card Authors](#)
- [Model Card Contact](#)

Dataset Card

- [Dataset Description](#)
 - [Dataset Summary](#)
 - [Supported Tasks and Leaderboards](#)
 - [Languages](#)
- [Dataset Structure](#)
- [Dataset Creation](#)
 - [Curation Rationale](#)
 - [Source Data](#)
 - [Annotations](#)
 - [Personal and Sensitive Information](#)
- [Considerations for Using the Data](#)
 - [Social Impact of Dataset](#)
 - [Discussion of Biases](#)
 - [Other Known Limitations](#)
- [Additional Information](#)
 - [Dataset Curators](#)
 - [Licensing Information](#)
 - [Citation Information](#)
 - [Contributions](#)

Model Details

Definition of Activity Fields

Activity fields are part qualifications for a job. In other words, they are groups of tasks and competencies necessary that together form the whole task domain of an occupation.

In an online job posting, activity fields occur mainly as phrases in the section where a posting lists the tasks and responsibilities of the job seeker. A phrase follows usually one of the following patterns:

- noun + verb (e.g. „Speisen zubereiten“)
- Verb + noun (e.g. „[Sie] kochen Speisen“)
- Noun + noun (e.g. „Zubereitung von Speisen“)
- Descriptive noun (e.g. „Speisenzubereitung“)

To define the concept of activity fields further, the following rules were applied to word sequences in job postings.

An activity field is matched if a sequence of words ...

- ... corresponds literally to the name of an activity field.
- ... is used synonymously with the name of an activity field.
- ... describes a requirement that is part of an activity field.
- ... describes a group of requirements that include an activity field.

An activity field is not matched if a sequence of words ...

- ... explicitly refers to the company of the job advertisement and not to the requirements for the applicant.
- ... neither explicitly nor implicitly refers to an activity field.

Uses

Direct Use

The model was built for research purposes in labor market analysis. Applying the model to a large set of job postings can generate insights into the prevalence and the demand for activity fields and inform experts and stakeholders.

The extraction models were trained on the free text fields of German online job postings. Accordingly, activity field entities can be extracted at the word sequence level. Furthermore, the extracted entities can be aggregated at the document level and used as a classification model.

Each model was trained on job descriptions for specific KLDB occupations. Therefore each model must be only used for job postings from this specific set. Activity fields are highly context-sensitive as a task in one occupation might mean something completely different in a different occupation (e.g., "Preparing pipes" relates to different tasks for a street builder compared to an industry mechanic).

Almost all models have some categories that have low performance. This is often due to the low occurrence rates of these activity fields. Interpreting any results derived from predictions of these categories should be treated carefully.

Out-of-Scope Use

The model should only be used on online job postings and not on other types of text: e.g., CVs or profiles of job applicants. The model is not trained for this task and would produce useless results.

Bias, Risks, and Limitations

The presented model has a very limited scope as it is only trained for extracting activity fields from job postings in a specific domain.

The data for the models were annotated by professionals from the Bertelsmann Foundation, which were instructed and supervised by an expert for activity fields. Nevertheless, the training and evaluation dataset might have some concept ambiguity and non-systematic concept errors.

The dependence on an occupation classification can further introduce bias into the predictions. As the standardized occupation is crucial in deciding which components to use for the predictions, any classification error will directly influence the usefulness of the activity field predictions. In the test dataset, around 10% of documents had to be discarded as they were incorrectly labeled with occupations.

Recommendations

It is imperative only to use the model as described above. Any occupation classification or normalizing that happens beforehand should be evaluated thoroughly.

Training Details

Training Data

This dataset is an annotated corpus of 15648 online job ads (OJA). The annotations describe activity fields found in the job ads. Corpora were annotated separately for the occupation groups and checked for consistency using Krippendorff's Alpha. For details, please see the dataset card.

Training Procedure

As described above, the model of each occupation consists of two components:

1. A rule-based span extractor searches a text for instances of certain entities in a text and extracts the position of the instances. The span extractor searches for direct matches of activity fields in texts but also searches the text via lexical attributes, linguistic features, operators, set affiliation, and operators. For example, the phrase “preparation of food” can be described as a pattern: “to prepare OR to cook,” followed by one or two optional tokens, followed by “food OR eating OR dishes.” In this way, different versions of the phrase can be found.
2. A statistical span extractor uses a deep learning model based on Convolutional-Neural-Network (CNN) and transition-based parsing. For the extraction of activity fields, the rule-based extractor is used to create a pre-annotation, and then the statistical model generalizes the extraction. Using both components together considerably improved both precision and recall.

When training the statistical model, a tok2vec layer is fine-tuned simultaneously. When combining the models into one unified pipeline, both components and their individual tok2vec layers are exported and added to the pipeline model.

Preprocessing

The rule-based and statistical models use the spaCy processing pipeline (e.g., tokenization, POS-tagging, etc.) For more information, see the [spaCy documentation](#).

Evaluation

Testing Data, Factors & Metrics

Testing Data

Each corpus was divided into training and validation data with a 20% split. See more details in the data set card.

Factors

The evaluation is conducted separately for each occupation domain. The domain evaluations also present results for each activity field individually.

Metrics

For the model, we report the micro-scores on word sequence (span) level: Precision, Recall, and F1-Score. We report Micro- and Macro-Scores on the document level for Precision, Recall, and F1. For each individual activity field, we also report the number of true positives, false positives, and false negatives.

The micro score is important to understand how well the models generalize the whole activity field domain. On the other hand, the macro-scores help to understand how well the model generalized over all individual activity fields. Stark differences in micro- and macro-scores can arise when there is, for example, one activity field that is relatively frequent and has a good individual score, but other smaller activity fields have a lower performance. This would lead to a high micro score and a low macro score.

For the purpose of analysis, the document-level metrics are more important and therefore given in more detail.

Results

We report results individually for each model. However, to get an overall impression of how well the approach works, we can give averages over all 27 models:

concept	p (avg)	r (avg)	f1 (avg)
Document Classification (Micro)	0.87	0.85	0.86
Document Classification (Macro)	0.82	0.77	0.78
Span Extraction & Classification	0.73	0.67	0.69

Results by model

Kaufmann/-frau für Büromanagement

Slug: buerokaufmann

This model is for occupations with the KldB-IDs ['71401', '71402', '71412', '71432', '71442', '71452']. The model has the following activity fields: ['HF1: Assistenz- und Sekretariatsaufgaben erledigen', 'HF2: Einkaufsprozesse planen und durchführen', 'HF3: Werbemaßnahmen umsetzen und Aufträge bearbeiten', 'HF4: Geschäftsvorgänge buchen', 'HF5: Geschäftsprozesse kontrollieren und steuern', 'HF6: Personalaufgaben bearbeiten']. On the word sequence level (Span Classification) the precision is 0.6, the recall is 0.5, and the f-score is 0.54. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.6	0.5	0.54
Classification (Micro)	0.81	0.81	0.81
Classification (Macro)	0.77	0.73	0.74

Splitted up by activity field the performance is the following. To test this model 103 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Assistenz- und Sekretariatsaufgaben erledigen	0.85	0.95	0.9	77	14	4
HF2: Einkaufsprozesse planen und durchführen	0.67	0.55	0.6	12	6	10
HF3: Werbemaßnahmen umsetzen und Aufträge bearbeiten	0.83	0.84	0.83	57	12	11
HF4: Geschäftsvorgänge buchen	0.8	0.8	0.8	32	8	8
HF5: Geschäftsprozesse kontrollieren und steuern	0.76	0.8	0.78	16	5	4
HF6: Personalaufgaben bearbeiten	0.69	0.45	0.55	9	4	11

KFZ-Mechatroniker/in

Slug: kfz

This model is for occupations with the KldB-IDs ['25201', '25212', '25222']. The model has the following activity fields: ['HF1: Standardisierte Service- und Wartungsarbeiten durchführen', 'HF2: Verschleißbehaftete mechanische und elektrische Systeme instand halten', 'HF3: Mechanische und elektrische Systeme montieren und demontieren', 'HF4: Mechatronische Systeme reparieren', 'HF5: Fahrzeugsysteme mit Expertensystemen diagnostizieren']. On the word sequence level (Span Classification) the precision is 0.88, the recall is 0.84, and the f-score is 0.86. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.88	0.84	0.86
Classification (Micro)	0.98	0.94	0.96
Classification (Macro)	0.98	0.93	0.95

Splitted up by activity field the performance is the following. To test this model 85 documents were used.

activity field	p	r	f1	tp	fp	fn
H1: Standardisierte Service- und Wartungsarbeiten durchführen	0.98	0.93	0.95	62	1	5
H2: Verschleißbehaftete mechanische und elektrische Systeme instand halten	0.98	0.96	0.97	64	1	3
H3: Mechanische und elektrische Systeme montieren und demontieren	0.99	0.94	0.96	67	1	4
H4: Mechatronische Systeme reparieren	0.98	0.97	0.97	57	1	2
H5: Fahrzeugsysteme mit Expertensystemen diagnostizieren	0.95	0.87	0.91	39	2	6

Anlagenmechaniker/-in für Sanitär-, Heizungs- und Klimatechnik

Slug: anlagenmechaniker

This model is for occupations with the KldB-IDs ['34201', '34202', '34212', '34222']. The model has the following activity fields: ['HF1: Rohre installieren', 'HF2: Raumluft-technische Anlagen erstellen', 'HF3: Heizungsanlagen erstellen', 'HF4: Sanitäre Anlagen erstellen', 'HF5: Thermische Solaranlagen erstellen']. On the word sequence level (Span Classification) the precision is 0.46, the recall is 0.67, and the f-score is 0.55. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.46	0.67	0.55
Classification (Micro)	0.71	0.9	0.79
Classification (Macro)	0.77	0.79	0.73

Splitted up by activity field the performance is the following. To test this model 118 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Rohre installieren	0.82	0.84	0.83	36	8	7
HF2: Raumluft-technische Anlagen erstellen	0.5	0.87	0.64	34	34	5
HF3: Heizungsanlagen erstellen	0.78	0.96	0.86	72	20	3
HF4: Sanitäre Anlagen erstellen	0.73	0.96	0.83	66	25	3
HF5: Thermische Solaranlagen erstellen	1	0.33	0.5	3	0	6

Friseur/in

Slug: **friseur**

This model is for occupations with the KldB-IDs ['82311', '82312']. The model has the following activity fields: ['HF1: Haare schneiden', 'HF2: Haare färben', 'HF3: Haare dauerhaft umformen', 'HF4: Haare föhnen und stylen', 'HF5: Gesicht dekorativ gestalten']. On the word sequence level (Span Classification) the precision is 0.83, the recall is 0.79, and the f-score is 0.81. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.83	0.79	0.81
Classification (Micro)	0.96	0.91	0.94
Classification (Macro)	0.9	0.87	0.88

Splitted up by activity field the performance is the following. To test this model 103 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Haare schneiden	1	0.91	0.95	53	0	5
HF2: Haare färben	0.98	0.89	0.94	51	1	6
HF3: Haare dauerhaft umformen	0.91	0.89	0.9	42	4	5
HF4: Haare föhnen und stylen	0.96	0.96	0.96	51	2	2
HF5: Gesicht dekorativ gestalten	0.67	0.67	0.67	2	1	1

Verkäufer/in & Einzelhandelskaufmann/-kauffrau

Slug: **verkauf**

This model is for occupations with the KldB-IDs ['62101', '62102', '62112', '62122', '62182']. The model has the following activity fields: ['HF1: Kassieren', 'HF2: Bedienen, beraten und verkaufen', 'HF3: Verkaufsfördernde Maßnahmen durchführen', 'HF4: Warenwirtschaftliche Prozesse durchführen', 'HF5: Im Kundenservice arbeiten', 'HF6: Operative Prozesse planen, kontrollieren und steuern', 'HF7: Strategische Prozesse planen, kontrollieren und steuern']. On the word sequence level (Span Classification) the precision is 0.79, the recall is 0.75, and the f-score is 0.77. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.79	0.75	0.77
Classification (Micro)	0.91	0.86	0.88
Classification (Macro)	0.77	0.66	0.69

Splitted up by activity field the performance is the following. To test this model 112 documents were used.

activity field	p	r	f1	tp	fp	fn
H1: Kassieren	0.91	0.97	0.94	60	6	2
H2: Bedienen, beraten und verkaufen	0.93	0.96	0.95	82	6	3
H3: Verkaufsfördernde Maßnahmen durchführen	1	0.78	0.88	29	0	8
H4: Warenwirtschaftliche Prozesse durchführen	0.95	0.92	0.94	61	3	5
H5: Im Kundenservice arbeiten	0.83	0.36	0.5	5	1	9
H6: Operative Prozesse planen, kontrollieren und steuern	0.5	0.23	0.32	3	3	10
H7: Strategische Prozesse planen, kontrollieren und steuern	0.29	0.4	0.33	2	5	3

Elektroniker/-in

Slug: **elektroniker**

This model is for occupations with the KldB-IDs ['26212', '26222', '26232', '26242', '26262', '26301']. The model has the following activity fields: ['HF1: Elektroinstallationen vor- und nachbereiten', 'HF2: Elektrische Betriebsmittel montieren und installieren', 'HF3: Elektrische Anlagen in Betrieb nehmen und instand halten', 'HF4: Steuerungen und Antriebssysteme montieren und installieren', 'HF5: Ortsveränderliche elektrische Geräte instand halten', 'HF6: Energietechnische Anlagen errichten und instand halten']. On the word sequence level (Span Classification) the precision is 0.66, the recall is 0.66, and the f-score is 0.66. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.66	0.66	0.66
Classification (Micro)	0.88	0.92	0.9
Classification (Macro)	0.85	0.82	0.82

Splitted up by activity field the performance is the following. To test this model 100 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Elektroinstallationen vor- und nachbereiten	0.93	0.95	0.94	77	6	4
HF2: Elektrische Betriebsmittel montieren und installieren	0.91	0.95	0.93	80	8	4
HF3: Elektrische Anlagen in Betrieb nehmen und instand halten	0.92	0.99	0.95	83	7	1
HF4: Steuerungen und Antriebssysteme montieren und installieren	0.73	0.7	0.71	16	6	7
HF5: Ortsveränderliche elektrische Geräte instand halten	1	0.56	0.71	5	0	4
HF6: Energietechnische Anlagen errichten und instand halten	0.62	0.79	0.7	15	9	4

Koch/Köchin

Slug: **koch**

This model is for occupations with the KldB-IDs ['29302', '29301', '29382']. The model has the following activity fields: ['H1: Spül- und Reinigungsarbeiten durchführen', 'H2: Salate und Vorspeisen vorbereiten und herstellen', 'H3: Beilagen und Suppen vorbereiten und herstellen', 'H4: Süßspeisen herstellen', 'H5: Fleisch- und Fischgerichte herstellen', 'H6: Küchenbetrieb leiten', 'H7: Im Service arbeiten']. On the word sequence level (Span Classification) the precision is 0.54, the recall is 0.77, and the f-score is 0.63. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.54	0.77	0.63
Classification (Micro)	0.76	0.88	0.82
Classification (Macro)	0.69	0.86	0.75

Splitted up by activity field the performance is the following. To test this model 106 documents were used.

activity field	p	r	f1	tp	fp	fn
A: Spül- und Reinigungsarbeiten durchführen	0.88	0.92	0.9	22	3	2
B: Salate und Vorspeisen vorbereiten und herstellen	0.79	0.9	0.84	46	12	5
C: Beilagen und Suppen vorbereiten und herstellen	0.8	0.87	0.84	41	10	6
D: Süßspeisen herstellen	0.77	0.85	0.81	40	12	7
E: Fleisch- und Fischgerichte herstellen	0.78	0.92	0.84	46	13	4
F: Küchenbetrieb leiten	0.33	1	0.5	4	8	0
G: Im Service arbeiten	0.5	0.56	0.53	5	5	4

Fachkraft Gastgewerbe

Slug: fkg

This model is for occupations with the KldB-IDs ['63302', '63301', '63312', '63322', '63382']. The model has the following activity fields: ['H1: Reinigungsarbeiten durchführen (Housekeeping)', 'H2: Veranstaltungen vorbereiten und im Warenlager arbeiten', 'H3: In der Küche helfen', 'H4: Im Service arbeiten', 'H5: An der Rezeption arbeiten', 'H6: Service leiten']. On the word sequence level (Span Classification) the precision is 0.64, the recall is 0.62, and the f-score is 0.63. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.64	0.62	0.63
Classification (Micro)	0.76	0.76	0.76
Classification (Macro)	0.64	0.66	0.6

Splitted up by activity field the performance is the following. To test this model 151 documents were used.

activity field	p	r	f1	tp	fp	fn
H1: Reinigungsarbeiten durchführen (Housekeeping)	0.41	1	0.58	11	16	0
H2: Veranstaltungen vorbereiten und im Warenlager arbeiten	0.67	0.23	0.34	6	3	20
H3: In der Küche helfen	0.9	0.51	0.66	19	2	18
H4: Im Service arbeiten	0.89	0.98	0.94	119	14	2
H5: An der Rezeption arbeiten	0.64	0.88	0.74	7	4	1
H6: Service leiten	0.32	0.37	0.34	7	15	12

Tiefbaufacharbeiter/in

Slug: tbf

This model is for occupations with the KldB-IDs ['32222', '32201', '32202', '32232', '32252']. The model has the following activity fields: ['H1: Vermessungs- und Erdbauarbeiten durchführen', 'H2: Schächte setzen, Rohrleitungen verlegen und instand setzen', 'H3: Pflasterflächen herstellen', 'H4: Verkehrsflächen instand setzen', 'H5: Einfache Hochbauarbeiten durchführen']. On the word sequence level (Span Classification) the precision is 0.7, the recall is 0.59, and the f-score is 0.64. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.7	0.59	0.64
Classification (Micro)	0.89	0.75	0.81
Classification (Macro)	0.87	0.7	0.77

Splitted up by activity field the performance is the following. To test this model 205 documents were used.

activity field	p	r	f1	tp	fp	fn
H1: Vermessungs- und Erdbauarbeiten durchführen	0.84	0.65	0.74	49	9	26
H2: Schächte setzen, Rohrleitungen verlegen und instand setzen	0.87	0.82	0.84	78	12	17
H3: Pflasterflächen herstellen	0.97	0.76	0.85	62	2	20
H4: Verkehrsflächen instand setzen	0.89	0.82	0.85	54	7	12
H5: Einfache Hochbauarbeiten durchführen	0.78	0.44	0.56	7	2	9

Hochbaufacharbeiter/in

Slug: **hbf**

This model is for occupations with the KldB-IDs ['32122', '32101', '32102', '32112', '32132', '32142', '32152', '32162', '32172']. The model has the following activity fields: ['H1: Einfache einschalige Baukörper erstellen', 'H2: Bauwerke erschließen', 'H3: Baukörper mit tragenden Funktionen herstellen', 'H4: Baukörper abdichten und dämmen', 'H5: Einfache Ausbauarbeiten durchführen', 'H6: Besondere Bauteile und Sichtmauerwerk herstellen']. On the word sequence level (Span Classification) the precision is 0.87, the recall is 0.57, and the f-score is 0.69. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.87	0.57	0.69
Classification (Micro)	0.9	0.71	0.79
Classification (Macro)	0.87	0.65	0.74

Splitted up by activity field the performance is the following. To test this model 142 documents were used.

activity field	p	r	f1	tp	fp	fn
H1: Einfache einschalige Baukörper erstellen	0.95	0.76	0.84	38	2	12
H2: Bauwerke erschließen	0.91	0.72	0.81	31	3	12
H3: Baukörper mit tragenden Funktionen herstellen	0.95	0.79	0.86	37	2	10
H4: Baukörper abdichten und dämmen	0.83	0.42	0.56	5	1	7
H5: Einfache Ausbauarbeiten durchführen	0.86	0.68	0.76	30	5	14
H6: Besondere Bauteile und Sichtmauerwerk herstellen	0.7	0.54	0.61	7	3	6

Ausbaufacharbeiter/in

Slug: **abf**

This model is for occupations with the KldB-IDs ['33112', '33101', '33102', '33122', '33132']. The model has the following activity fields: ['H1: Wandfliesen im Dünnbett verlegen', 'H2: Bodenfliesen im Dünnbett verlegen', 'H3: Zubehör und zusätzliche Materialien einbauen', 'H4: Spezielle Fliesen- und Plattenverlegearbeiten durchführen', 'H5: Dämm-, Trocken-, Putz- und Estricharbeiten ausführen', 'H6: Einfache Hochbauarbeiten durchführen']. On the word sequence level (Span Classification) the precision is 0.74, the recall is 0.53, and the f-score is 0.62. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.74	0.53	0.62
Classification (Micro)	0.83	0.74	0.78
Classification (Macro)	0.87	0.7	0.76

Splitted up by activity field the performance is the following. To test this model 112 documents were used.

activity field	p	r	f1	tp	fp	fn
H1: Wandfliesen im Dünnbett verlegen	0.8	0.77	0.79	37	9	11
H2: Bodenfliesen im Dünnbett verlegen	0.83	0.78	0.81	40	8	11
H3: Zubehör und zusätzliche Materialien einbauen	0.79	0.71	0.75	34	9	14
H4: Spezielle Fliesen- und Plattenverlegearbeiten durchführen	0.76	0.74	0.75	32	10	11
H5: Dämm-, Trocken-, Putz- und Estricharbeiten ausführen	1	0.68	0.81	26	0	12
H6: Einfache Hochbauarbeiten durchführen	1	0.5	0.67	1	0	1

Fachinformatiker/-in (3 Fachrichtungen – Systemintegration, Anwendungstechnik und digitale Vernetzung)

Slug: informatiker

This model is for occupations with the KldB-IDs ['43102', '43112', '43122', '43152', '43412']. The model has the following activity fields: ['HF1: Einfache IT-Systeme zusammenbauen, installieren und warten (A)', 'HF2: IT-Netzwerke installieren, integrieren und konfigurieren (B)', 'HF3: IT-Sicherheit anwenden, Serverdienste installieren und warten (C)', 'HF4: Mit Datenbanken arbeiten (D1, D2)', 'HF5: Schnittstellen programmieren (D3)', 'HF6: Anwendungen programmieren (E1, E2, F2)', 'HF7: Cyber-physischer-Systeme einrichten (E3)', 'HF8: Systeme automatisieren und pflegen (F1)', 'HF9: Cyber-physischer-Systeme optimieren (F3)']. On the word sequence level (Span Classification) the precision is 0.78, the recall is 0.56, and the f-score is 0.65. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.78	0.56	0.65
Classification (Micro)	0.87	0.87	0.87
Classification (Macro)	0.67	0.64	0.65

Splitted up by activity field the performance is the following. To test this model 109 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Einfache IT-Systeme zusammenbauen, installieren und warten (A)	0.91	0.9	0.91	75	7	8
HF2: IT-Netzwerke installieren, integrieren und konfigurieren (B)	0.94	0.94	0.94	78	5	5
HF3: IT-Sicherheit anwenden, Serverdienste installieren und warten (C)	0.98	0.97	0.97	86	2	3
HF4: Mit Datenbanken arbeiten (D1, D2)	0.81	0.71	0.76	25	6	10
HF5: Schnittstellen programmieren (D3)	0.77	0.77	0.77	10	3	3
HF6: Anwendungen programmieren (E1, E2, F2)	0.59	0.79	0.67	34	24	9
HF7: Cyber-physischer-Systeme einrichten (E3)	0	0	0	0	0	2
HF8: Systeme automatisieren und pflegen (F1)	1	0.68	0.81	13	0	6
HF9: Cyber-physischer-Systeme optimieren (F3)	0	0	0	0	0	0

Fachkraft Metalltechnik – Fachrichtung Konstruktionstechnik

Slug: metall_konstruktion

This model is for occupations with the KldB-IDs ['24411', '24412']. The model has the following activity fields: ['HF1: Bauteile mit handgeführten Werkzeugen und Maschinen herstellen', 'HF2: Einzelteile mit Werkzeugmaschinen drehen und fräsen', 'HF3: Baugruppen und Metallkonstruktionen montieren und demontieren', 'HF4: Bleche und Rohre umformen und trennen', 'HF5: Bauteile und Baugruppen schweißen']. On the word sequence level (Span Classification) the precision is 0.8, the recall is 0.64, and the f-score is 0.71. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.8	0.64	0.71
Classification (Micro)	0.86	0.81	0.84
Classification (Macro)	0.87	0.68	0.69

Splitted up by activity field the performance is the following. To test this model 97 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Bauteile mit handgeführten Werkzeugen und Maschinen herstellen	0.82	0.7	0.75	23	5	10
HF2: Einzelteile mit Werkzeugmaschinen drehen und fräsen	1	0.08	0.15	1	0	11
HF3: Baugruppen und Metallkonstruktionen montieren und demontieren	0.92	0.94	0.93	61	5	4
HF4: Bleche und Rohre umformen und trennen	0.66	0.7	0.68	19	10	8
HF5: Bauteile und Baugruppen schweißen	0.93	0.96	0.95	43	3	2

Fachkraft Metalltechnik – Fachrichtung Zerspanungstechnik

Slug: metall_zerspanung

This model is for occupations with the KldB-IDs ['24232']. The model has the following activity fields: ['HF1: Bauteile mit handgeführten Werkzeugen und Maschinen herstellen', 'HF2: Einzelteile mit Werkzeugmaschinen drehen und fräsen', 'HF3: Baugruppen und Metallkonstruktionen montieren und demontieren', 'HF4: CNC-Drehmaschinen und -Drehmaschinensysteme bedienen, überwachen und warten', 'HF5: CNC-Fräsmaschinen und -Fräsmaschinensysteme bedienen, überwachen und warten']. On the word sequence level (Span Classification) the precision is 0.74, the recall is 0.63, and the f-score is 0.68. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.74	0.63	0.68
Classification (Micro)	0.9	0.89	0.89
Classification (Macro)	0.67	0.73	0.69

Splitted up by activity field the performance is the following. To test this model 107 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Bauteile mit handgeführten Werkzeugen und Maschinen herstellen	0.3	0.4	0.34	6	14	9
HF2: Einzelteile mit Werkzeugmaschinen drehen und fräsen	0.91	0.88	0.89	77	8	11
HF3: Baugruppen und Metallkonstruktionen montieren und demontieren	0.2	0.5	0.29	1	4	1
HF4: CNC-Drehmaschinen und -Drehmaschinensysteme bedienen, überwachen und warten	0.99	0.95	0.97	92	1	5
HF5: CNC-Fräsmaschinen und -Fräsmaschinensysteme bedienen, überwachen und warten	0.98	0.92	0.95	89	2	8

Groß- und Außenhandelsmanagement

Slug: grosshandel

This model is for occupations with the KldB-IDs ['61212']. The model has the following activity fields: ['HF1: Waren beschaffen ', 'HF2: Konten führen und Zahlungsverkehr abwickeln', 'HF3: Kunden beraten und Waren verkaufen', 'HF4: Im Großhandelslager arbeiten', 'HF5: Waren im Außenhandel exportieren']. On the word sequence level (Span Classification) the precision is 0.61, the recall is 0.57, and the f-score is 0.59. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.61	0.57	0.59
Classification (Micro)	0.84	0.88	0.86
Classification (Macro)	0.83	0.87	0.85

Splitted up by activity field the performance is the following. To test this model 116 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Waren beschaffen	0.81	0.79	0.8	58	14	15
HF2: Konten führen und Zahlungsverkehr abwickeln	0.91	0.77	0.83	48	5	14
HF3: Kunden beraten und Waren verkaufen	0.91	0.98	0.95	104	10	2
HF4: Im Großhandelslager arbeiten	0.9	0.89	0.89	71	8	9
HF5: Waren im Außenhandel exportieren	0.64	0.94	0.76	45	25	3

Industriemechaniker/in

Slug: industriemechaniker

This model is for occupations with the KldB-IDs ['25101', '25102', '25112']. The model has the following activity fields: ['HF1: Werkstücke durch manuelle und maschinelle Zerspanung herstellen und bearbeiten', 'HF2: Mechanische Verbindungen herstellen', 'HF3: Bauteile auf numerisch gesteuerten Werkzeugmaschinen herstellen', 'HF4: Steuerungen der Fluidtechnik aufbauen, prüfen und in Betrieb nehmen', 'HF5: Elektrotechnische Systeme installieren, prüfen und in Betrieb nehmen', 'HF6: Technische Systeme in Stand halten']. On the word sequence level (Span Classification) the precision is 0.78, the recall is 0.59, and the f-score is 0.67. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.78	0.59	0.67
Classification (Micro)	0.88	0.77	0.82
Classification (Macro)	0.87	0.68	0.73

Splitted up by activity field the performance is the following. To test this model 88 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Werkstücke durch manuelle und maschinelle Zer-spanung herstellen und bearbeiten	0.93	0.59	0.72	13	1	9
HF2: Mechanische Verbindungen herstellen	0.84	0.95	0.89	54	10	3
HF3: Bauteile auf numerisch gesteuerten Werkzeugmaschinen herstellen	1	0.38	0.55	9	0	15
HF4: Steuerungen der Fluidtechnik aufbauen, prüfen und in Betrieb nehmen	0.92	0.57	0.71	12	1	9
HF5: Elektrotechnische Systeme installieren, prüfen und in Betrieb nehmen	0.55	0.67	0.6	6	5	3
HF6: Technische Systeme in Stand halten	0.96	0.91	0.94	51	2	5

Industrieelektriker/in

Slug: **industrieelektriker**

This model is for occupations with the KldB-IDs ['26252']. The model has the following activity fields: ['HF1: Komponenten und Betriebsmittel bearbeiten und montieren', 'HF2: Elektrische Anlagen und Maschinen installieren', 'HF3: Elektrische Anlagen, Maschinen und Geräte instand setzen', 'HF4: Elektrische Anlagen, Maschinen und Geräte in Betrieb nehmen']. On the word sequence level (Span Classification) the precision is 0.85, the recall is 0.79, and the f-score is 0.82. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.85	0.79	0.82
Classification (Micro)	0.95	0.92	0.93
Classification (Macro)	0.95	0.91	0.93

Splitted up by activity field the performance is the following. To test this model 111 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Komponenten und Betriebsmittel bearbeiten und montieren	0.92	0.89	0.9	47	4	6
HF2: Elektrische Anlagen und Maschinen installieren	0.95	0.92	0.93	54	3	5
HF3: Elektrische Anlagen, Maschinen und Geräte instand setzen	0.96	0.99	0.97	76	3	1
HF4: Elektrische Anlagen, Maschinen und Geräte in Betrieb nehmen	0.98	0.85	0.91	50	1	9

Fachlagerist/in

Slug: **lager**

This model is for occupations with the KldB-IDs ['51311', '51312']. The model has the following activity fields: ['HF1: Waren annehmen', 'HF2: Fördermittel einsetzen', 'HF3: Waren lagern', 'HF4: Waren kommissionieren', 'HF5: Waren versenden', 'HF6: Logistische Prozesse planen und optimieren', 'HF7: Waren beschaffen']. On the word sequence level (Span Classification) the precision is 0.86, the recall is 0.73, and the f-score is 0.79. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.86	0.73	0.79
Classification (Micro)	0.96	0.85	0.9
Classification (Macro)	0.82	0.68	0.72

Splitted up by activity field the performance is the following. To test this model 107 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Waren annehmen	0.98	0.89	0.93	49	1	6
HF2: Fördermittel einsetzen	0.92	0.97	0.94	56	5	2
HF3: Waren lagern	1	0.51	0.68	22	0	21
HF4: Waren kommissionieren	0.99	0.97	0.98	66	1	2
HF5: Waren versenden	0.95	0.95	0.95	73	4	4
HF6: Logistische Prozesse planen und optimieren	0.89	0.44	0.59	8	1	10
HF7: Waren beschaffen	0	0	0	0	0	3

Gärtner:in

Slug: **gaertner**

This model is for occupations with the KldB-IDs ['12101', '12102', '12142']. The model has the following activity fields: ['HF1: Erdarbeiten ausführen und Maschinen bedienen', 'HF2: Befestigte Flächen herstellen und Entwässerungssysteme einbauen', 'HF3: Bauwerke herstellen', 'HF4: Vegetationstechnische Arbeiten und Neupflanzungen durchführen', 'HF5: Gärtnerische Pflegemaßnahmen durchführen']. On the word sequence level (Span Classification) the precision is 0.76, the recall is 0.73, and the f-score is 0.75. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.76	0.73	0.75
Classification (Micro)	0.87	0.8	0.83
Classification (Macro)	0.84	0.73	0.78

Splitted up by activity field the performance is the following. To test this model 91 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Erdarbeiten ausführen und Maschinen bedienen	0.81	0.67	0.73	22	5	11
HF2: Befestigte Flächen herstellen und Entwässerungssysteme einbauen	0.92	0.81	0.86	22	2	5
HF3: Bauwerke herstellen	0.71	0.6	0.65	12	5	8
HF4: Vegetationstechnische Arbeiten und Neupflanzungen durchführen	0.83	0.6	0.7	15	3	10
HF5: Gärtnerische Pflegemaßnahmen durchführen	0.92	0.97	0.95	71	6	2

Tischler:in

Slug: **tischler**

This model is for occupations with the KldB-IDs ['22301', '22302', '22342']. The model has the following activity fields: ['HF1: Produkte an Holzbearbeitungsmaschinen herstellen', 'HF2: Einfache Produkte in Einzelfertigung herstellen und zusammenbauen', 'HF3: Oberflächen beschichten, lackieren und veredeln', 'HF4: Möbel demontieren und montieren', 'HF5: Bauelemente demontieren und montieren', 'HF6: Wartungs- und Instandhaltungsarbeiten durchführen']. On the word sequence level (Span Classification) the precision is 0.64, the recall is 0.66, and the f-score is 0.65. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.64	0.66	0.65
Classification (Micro)	0.79	0.86	0.82
Classification (Macro)	0.79	0.82	0.79

Splitted up by activity field the performance is the following. To test this model 52 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Produkte an Holzbearbeitungsmaschinen herstellen	0.9	0.96	0.93	27	3	1
HF2: Einfache Produkte in Einzelfertigung herstellen und zusammenbauen	0.88	0.64	0.74	14	2	8
HF3: Oberflächen beschichten, lackieren und veredeln	0.83	0.76	0.79	19	4	6
HF4: Möbel demontieren und montieren	0.84	0.97	0.9	31	6	1
HF5: Bauelemente demontieren und montieren	0.62	0.92	0.74	23	14	2
HF6: Wartungs- und Instandhaltungsarbeiten durchführen	0.67	0.67	0.67	4	2	2

Bäcker:in

Slug: baecker

This model is for occupations with the KldB-IDs ['29222']. The model has the following activity fields: ['HF1: Teige herstellen', 'HF2: Teige zu Brot und Kleingebäck aufarbeiten', 'HF3: Feine Backwaren aus Teigen aufarbeiten', 'HF4: Feine Backwaren aus Massen herstellen, Torten herstellen', 'HF5: Backwarensnacks und kleine Gerichte herstellen', 'HF6: Brot, Kleingebäck und Feine Backwaren abbacken']. On the word sequence level (Span Classification) the precision is 0.62, the recall is 0.72, and the f-score is 0.67. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.62	0.72	0.67
Classification (Micro)	0.82	0.9	0.86
Classification (Macro)	0.82	0.89	0.85

Splitted up by activity field the performance is the following. To test this model 78 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Teige herstellen	0.91	0.94	0.92	49	5	3
HF2: Teige zu Brot und Kleingebäck aufarbeiten	0.9	0.83	0.86	38	4	8
HF3: Feine Backwaren aus Teigen aufarbeiten	0.78	0.83	0.8	35	10	7
HF4: Feine Backwaren aus Massen herstellen, Torten herstellen	0.78	0.95	0.86	40	11	2
HF5: Backwarensnacks und kleine Gerichte herstellen	0.72	0.85	0.78	29	11	5
HF6: Brot, Kleingebäck und Feine Backwaren abbacken	0.82	0.94	0.88	50	11	3

Fachkraft Lebensmittelhandwerk (Bäckerei)

Slug: fk_baeckerei

This model is for occupations with the KldB-IDs ['62312', '62301']. The model has the following activity fields: ['HF1: Backwaren und Betriebsmittel annehmen, prüfen und lagern', 'HF2: Snacks, kleine Gerichte und Getränke herstellen', 'HF3: Teiglinge zu Backwaren abbacken', 'HF4: Kunden beraten, Waren verkaufen', 'HF5: Kasservorgänge durchführen', 'HF6: Verkaufsfördernde Maßnahmen durchführen']. On the word sequence level (Span Classification) the precision is 0.87, the recall is 0.84, and the f-score is 0.86. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.87	0.84	0.86
Classification (Micro)	0.97	0.9	0.93
Classification (Macro)	0.97	0.89	0.93

Splitted up by activity field the performance is the following. To test this model 101 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Backwaren und Betriebsmittel annehmen, prüfen und lagern	0.98	0.92	0.95	58	1	5
HF2: Snacks, kleine Gerichte und Getränke herstellen	1	0.81	0.9	44	0	10
HF3: Teiglinge zu Backwaren abbacken	0.91	0.91	0.91	39	4	4
HF4: Kunden beraten, Waren verkaufen	1	0.99	0.99	85	0	1
HF5: Kasservorgänge durchführen	0.95	0.89	0.92	42	2	5
HF6: Verkaufsfördernde Maßnahmen durchführen	0.96	0.81	0.88	47	2	11

Maler:in

Slug: maler

This model is for occupations with the KldB-IDs ['33211', '33212']. The model has the following activity fields: ['HF1: Innen und außen anstreichen', 'HF2: Objekte lackieren', 'HF3: Putzarbeiten durchführen', 'HF4: Trockenbauarbeiten durchführen', 'HF5: Tapeten und Bodenbeläge kleben', 'HF6: Dämmarbeiten durchführen']. On the word sequence level (Span Classification) the precision is 0.79, the recall is 0.8, and the f-score is 0.79. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.79	0.8	0.79
Classification (Micro)	0.91	0.92	0.92
Classification (Macro)	0.88	0.92	0.9

Splitted up by activity field the performance is the following. To test this model 103 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Innen und außen anstreichen	0.94	0.93	0.94	80	5	6
HF2: Objekte lackieren	0.96	0.94	0.95	77	3	5
HF3: Putzarbeiten durchführen	0.91	0.89	0.9	39	4	5
HF4: Trockenbauarbeiten durchführen	0.88	0.89	0.88	50	7	6
HF5: Tapeten und Bodenbeläge kleben	0.88	0.96	0.92	68	9	3
HF6: Dämmarbeiten durchführen	0.73	0.89	0.8	16	6	2

Gebäudereiniger:in

Slug: reiniger

This model is for occupations with the KldB-IDs ['54101', '54112', '54132', '54142', '54152', '54182']. The model has the following activity fields: ['HF1: Unterhaltsreinigung durchführen', 'HF2: Glasflächen, Fenster und Rahmen reinigen', 'HF3: Desinfizierende Reinigung durchführen ', 'HF4: Fassaden- und Industriereinigung durchführen', 'HF5: Sonderreinigungen durchführen']. On the word sequence level (Span Classification) the precision is 0.78, the recall is 0.79, and the f-score is 0.79. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.78	0.79	0.79
Classification (Micro)	0.89	0.86	0.88
Classification (Macro)	0.78	0.75	0.76

Splitted up by activity field the performance is the following. To test this model 104 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Unterhaltsreinigung durchführen	0.98	0.94	0.96	79	2	5
HF2: Glasflächen, Fenster und Rahmen reinigen	1	0.78	0.88	7	0	2
HF3: Desinfizierende Reinigung durchführen	0.6	0.75	0.67	3	2	1
HF4: Fassaden- und Industriereinigung durchführen	0.53	0.57	0.55	8	7	6
HF5: Sonderreinigungen durchführen	0.8	0.73	0.76	8	2	3

Berufskraftfahrer:in

Slug: **kraftfahrer**

This model is for occupations with the KldB-IDs ['52122', '52132']. The model has the following activity fields: ['HF1: Touren planen und dokumentieren', 'HF2: Fahrzeug vorbereiten', 'HF3: Güter befördern', 'HF4: Fahrzeug führen', 'HF5: Fahrzeug warten und pflegen', 'HF6: Personen befördern']. On the word sequence level (Span Classification) the precision is 0.81, the recall is 0.71, and the f-score is 0.76. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.81	0.71	0.76
Classification (Micro)	0.92	0.86	0.89
Classification (Macro)	0.91	0.79	0.83

Splitted up by activity field the performance is the following. To test this model 117 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Touren planen und dokumentieren	1	0.42	0.59	10	0	14
HF2: Fahrzeug vorbereiten	0.87	0.84	0.85	61	9	12
HF3: Güter befördern	0.9	0.93	0.91	78	9	6
HF4: Fahrzeug führen	1	0.98	0.99	112	0	2
HF5: Fahrzeug warten und pflegen	0.79	0.67	0.72	22	6	11
HF6: Personen befördern	0.9	0.9	0.9	9	1	1

Kaufmann/Kauffrau für Spedition und Logistikdienstleistung

Slug: **spedition**

This model is for occupations with the KldB-IDs ['51622']. The model has the following activity fields: ['HF1: Logistische Leistungen anbieten', 'HF2: Frachtaufträge im Güterkraftverkehr disponieren', 'HF3: Speditionsaufträge im Sammelgut- und Systemverkehr bearbeiten', 'HF4: Luft- und Seefrachtaufträge bearbeiten', 'HF5: Außenhandelsaufträge abwickeln', 'HF6: Geschäftsvorgänge steuern und kontrollieren']. On the word sequence level (Span Classification) the precision is 0.62, the recall is 0.56, and the f-score is 0.59. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.62	0.56	0.59
Classification (Micro)	0.88	0.84	0.86
Classification (Macro)	0.87	0.84	0.85

Splitted up by activity field the performance is the following. To test this model 113 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Logistische Leistungen anbieten	0.92	0.9	0.91	87	8	10
HF2: Frachtaufträge im Güterkraftverkehr disponieren	0.84	0.75	0.79	48	9	16
HF3: Speditionsaufträge im Sammelgut- und Systemverkehr bearbeiten	0.86	0.74	0.79	48	8	17
HF4: Luft- und Seefrachtaufträge bearbeiten	0.87	0.87	0.87	46	7	7
HF5: Außenhandelsaufträge abwickeln	0.92	0.93	0.93	57	5	4
HF6: Geschäftsvorgänge steuern und kontrollieren	0.85	0.83	0.84	55	10	11

Slug: sicherheit

This model is for occupations with the KldB-IDs ['53111', '53112']. The model has the following activity fields: ['HF1: Menschen und Objekte schützen', 'HF2: Schutz- und Sicherungsmaßnahmen im öffentlichen (Verkehrs)Raum durchführen', 'HF3: Veranstaltungsdienst durchführen', 'HF4: Sicherheitstechnische Einrichtungen und Hilfsmittel einsetzen und anwenden', 'HF5: Sicherheitsrelevante Sachverhalte ermitteln, aufklären und dokumentieren', 'HF6: Sicherheitskonzepte erstellen und Geschäftsprozesse durchführen']. On the word sequence level (Span Classification) the precision is 0.71, the recall is 0.59, and the f-score is 0.64. On the document level the model has the following performance:

concept	p	r	f1
Span Extraction (Micro)	0.71	0.59	0.64
Classification (Micro)	0.82	0.89	0.85
Classification (Macro)	0.72	0.73	0.71

Splitted up by activity field the performance is the following. To test this model 97 documents were used.

activity field	p	r	f1	tp	fp	fn
HF1: Menschen und Objekte schützen	0.92	0.98	0.95	84	7	2
HF2: Schutz- und Sicherungsmaßnahmen im öffentlichen (Verkehrs)Raum durchführen	1	0.56	0.71	5	0	4
HF3: Veranstaltungsdienst durchführen	0.86	1	0.92	6	1	0
HF4: Sicherheitstechnische Einrichtungen und Hilfsmittel einsetzen und anwenden	0.72	0.91	0.81	29	11	3
HF5: Sicherheitsrelevante Sachverhalte ermitteln, aufklären und dokumentieren	0.12	0.17	0.14	1	7	5
HF6: Sicherheitskonzepte erstellen und Geschäftsprozesse durchführen	0.67	0.75	0.71	6	3	2

Model Examination

A prerequisite for the targeted extraction of activity fields is the restriction to target occupations for which the activity fields were developed. In theory, this constraint is detached from the performance of the model. In practice, a poor occupation classification model leads to downstream poor results of the activity field extraction.

The extraction of activity fields works well using a hybrid approach in which a rule-based approach is supplemented by a statistical model. Rule-based models initially have a high precision because word sequence patterns can be formulated precisely. The weakness of rule-based models is that they cannot generalize beyond the predefined patterns. However, this weakness can be partially compensated by the statistical model, which can detect unseen word sequences via the semantic structure of the training examples.

The analysis shows that the definition of the activity fields - and their proximity to the actual activity in the respective occupation - greatly influence the quality of the models. This is particularly evident in the case of the model for the activity field "hospitality and service": hospitality or customer service skill jobs can be advertised in different areas, each with a very different context and activities. To mitigate this challenge, the number of documents from different contexts would have to be weighted (in the case of the hospitality specialist, e.g. hotel reception, hotel cleaning, kitchen help, restaurant service) or significantly more documents would have to be annotated altogether. The evaluation of the models shows that activity fields are named in job advertisements, and their extraction is possible with good to excellent results.

Environmental Impact

Carbon emissions can be estimated using the [Machine Learning Impact calculator](#) presented in [Lacoste et al. \(2019\)](#). As the model was trained only locally and the hardware requirements are low for a small amount of data, no carbon emissions were calculated.

- **Hardware Type:** Local Machine; M1
- **Hours used:** 24h
- **Cloud Provider:** -
- **Compute Region:** -
- **Carbon Emitted:** -

Software

python = “^3.8” spacy = “^3.3.0”

Model Card Authors

Johannes Müller (&effect data solutions GmbH)

Model Card Contact

Johannes Müller (johannes.mueller@and-effect.com)

Dataset Card: Goldstandard for Activity Fields in German Online Job Ads

Dataset Summary

This dataset is an annotated corpus of 15648 online job ads (OJA). The annotations describe **activity fields** found in the job ads. Activity fields (German: “Teilqualifikationen”) are subsets of skills or competencies which together form the full qualifications gained by a two- or three-year vocational training (“Berufsausbildung”).

They are specific to occupations and are linked to the German occupation taxonomy KldB-2010. For each occupation, there are between 4 and 8 activity fields, forming an “activity field model”.

In the presented dataset, there are 27 activity field models. The dataset is split up by occupation. Use the meta.json file to get more information in the activity field models and link them to specific occupation KldB-IDs.

You can download the data in .json or .spacy DocBin-binary format.

Each corpus was divided into training and validation data with a 20% split.

Supported Tasks and Leaderboards

Annotation provided both as span annotations for named-entity-recognition and/or span categorization tasks under doc["spans"]. Annotations are also provided as categories under doc["cats"] for text classification tasks.

Note: Many activity field models have overlapping span annotations. That is because many competences hint towards multiple activity fields.

Languages

The dataset features only German Online Job Ads.

Dataset Structure

The dataset is provided in both .json format and Spacy's [DocBin format](#) (.spacy).

The .json data has the following schema:

```
[  
  {  
    "text": str # Full text of the job ad,  
    "spans": [  
      {  
        "start_char": int, # First character position of span  
        "stop_char": int, # Last character position of span  
        "label": str, # Name of the activity field  
        "kb_id": uuid # This is a unique UUID  
      },  
      ...  
    ]  
  },  
  {"cats": {  
    "--label--": 1, # Each activity field found in the online job ad as a key-value pair of "label": 1  
    "--label--": 1,  
    ...  
  }}]
```

Activity Field Model Definition and Annotation Quality

Kaufmann/-frau für Büromanagement

Slug: buerokaufmann

This model is for occupations with the KldB-IDs ['71401', '71402', '71412', '71432', '71442', '71452']. The model has the following activity fields: ['HF1: Assistenz- und Sekretariatsaufgaben erledigen', 'HF2: Einkaufsprozesse planen und durchführen', 'HF3: Werbemaßnahmen umsetzen und Aufträge bearbeiten', 'HF4: Geschäftsvorgänge buchen', 'HF5: Geschäftsprozesse kontrollieren und steuern', 'HF6: Personalaufgaben bearbeiten']. For the goldstandard 517 documents were annotated by 2 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Assistenz- und Sekretariatsaufgaben erledigen	0.738943	424
1 HF2: Einkaufsprozesse planen und durchführen	0.666273	114
2 HF3: Werbemaßnahmen umsetzen und Aufträge bearbeiten	0.768199	298
3 HF4: Geschäftsvorgänge buchen	0.817661	204
4 HF5: Geschäftsprozesse kontrollieren und steuern	0.563964	120
5 HF6: Personalaufgaben bearbeiten	0.667336	102

KFZ-Mechatroniker/in

Slug: kfz

This model is for occupations with the KldB-IDs ['25201', '25212', '25222']. The model has the following activity fields: ['HF1: Standardisierte Service- und Wartungsarbeiten durchführen', 'HF2: Verschleißbehaftete mechanische und elektrische Systeme instand halten', 'HF3: Mechanische und elektrische Systeme montieren und demontieren', 'HF4: Mechatronische Systeme reparieren', 'HF5: Fahrzeugsysteme mit Expertensystemen diagnostizieren']. For the goldstandard 425 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Standardisierte Service- und Wartungsarbeiten durchführen	0.892	308
1 HF2: Verschleißbehaftete mechanische und elektrische Systeme instand halten	0.73	318
2 HF3: Mechanische und elektrische Systeme montieren und demontieren	0.808	332
3 HF4: Mechatronische Systeme reparieren	0.764	286
4 HF5: Fahrzeugsysteme mit Expertensystemen diagnostizieren	0.829	213

Anlagenmechaniker/-in für Sanitär-, Heizungs- und Klimatechnik

Slug: anlagenmechaniker

This model is for occupations with the KldB-IDs ['34201', '34202', '34212', '34222', '34232']. The model has the following activity fields: ['HF1: Rohre installieren', 'HF2: Raumluft-technische Anlagen erstellen', 'HF3: Heizungsanlagen erstellen', 'HF4: Sanitäre Anlagen erstellen', 'HF5: Thermische Solaranlagen erstellen']. For the goldstandard 593 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Rohre installieren	0.819	215
1 HF2: Raumluft-technische Anlagen erstellen	0.813	153
2 HF3: Heizungsanlagen erstellen	0.545	362
3 HF4: Sanitäre Anlagen erstellen	0.702	360
4 HF5: Thermische Solaranlagen erstellen	0.697	35

Friseur/in

Slug: friseur

This model is for occupations with the KldB-IDs ['82311', '82312']. The model has the following activity fields: ['HF1: Haare schneiden', 'HF2: Haare färben', 'HF3: Haare dauerhaft umformen', 'HF4: Haare föhnen und stylen', 'HF5: Gesicht dekorativ gestalten']. For the goldstandard 518 documents were annotated by 2 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Haare schneiden	0.705882	299
1 HF2: Haare färben	0.462486	295
2 HF3: Haare dauerhaft umformen	0.749216	236
3 HF4: Haare föhnen und stylen	0.765625	268
4 HF5: Gesicht dekorativ gestalten	0.820829	28

Verkäufer/in & Einzelhandelskaufmann/-kauffrau

Slug: verkauf

This model is for occupations with the KldB-IDs ['62101', '62102', '62112', '62122', '62182']. The model has the following activity fields: ['HF1: Kassieren', 'HF2: Bedienen, beraten und verkaufen', 'HF3: Verkaufsfördernde Maßnahmen durchführen', 'HF4: Warenwirtschaftliche Prozesse durchführen', 'HF5: Im Kundenservice arbeiten', 'HF6: Operative Prozesse planen, kontrollieren und steuern', 'HF7: Strategische Prozesse planen, kontrollieren und steuern']. For the goldstandard 564 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Kassieren	0.938	328
1 HF2: Bedienen, beraten und verkaufen	0.771	429
2 HF3: Verkaufsfördernde Maßnahmen durchführen	0.842	214
3 HF4: Warenwirtschaftliche Prozesse durchführen	0.874	361
4 HF5: Im Kundenservice arbeiten	0.494	72
5 HF6: Operative Prozesse planen, kontrollieren und steuern	0.515	83
6 HF7: Strategische Prozesse planen, kontrollieren und steuern	0.528	26

Elektroniker/-in

Slug: elektroniker

This model is for occupations with the KldB-IDs ['26212', '26222', '26232', '26242', '26252', '26262']. The model has the following activity fields: ['HF1: Elektroinstallationen vor- und nachbereiten', 'HF2: Elektrische Betriebsmittel montieren und installieren', 'HF3: Elektrische Anlagen in Betrieb nehmen und instand halten', 'HF4: Steuerungen und Antriebssysteme montieren und installieren', 'HF5: Ortsveränderliche elektrische Geräte instand halten', 'HF6: Energietechnische Anlagen errichten und instand halten']. For the goldstandard 503 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Elektroinstallationen vor- und nachbereiten	0.356	365
1 HF2: Elektrische Betriebsmittel montieren und installieren	0.345	383
2 HF3: Elektrische Anlagen in Betrieb nehmen und instand halten	0.334	404
3 HF4: Steuerungen und Antriebssysteme montieren und installieren	0.637	156
4 HF5: Ortsveränderliche elektrische Geräte instand halten	0.557	40
5 HF6: Energietechnische Anlagen errichten und instand halten	0.369	83

Koch/Köchin

Slug: koch

This model is for occupations with the KldB-IDs ['29302', '29301', '29382']. The model has the following activity fields: ['H1: Spül- und Reinigungsarbeiten durchführen', 'H2: Salate und Vorspeisen vorbereiten und herstellen', 'H3: Beilagen und Suppen vorbereiten und herstellen', 'H4: Süßspeisen herstellen', 'H5: Fleisch- und Fischgerichte herstellen', 'H6: Küchenbetrieb leiten', 'H7: Im Service arbeiten']. For the goldstandard 969 documents were annotated by 6 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 H1: Spül- und Reinigungsarbeiten durchführen	0.74	300
1 H2: Salate und Vorspeisen vorbereiten und herstellen	0.7	550
2 H3: Beilagen und Suppen vorbereiten und herstellen	0.69	501
3 H4: Süßspeisen herstellen	0.61	423
4 H5: Fleisch- und Fischgerichte herstellen	0.52	450
5 H6: Küchenbetrieb leiten	0.27	74
6 H7: Im Service arbeiten	0.14	73

Fachkraft Gastgewerbe

Slug: fkg

This model is for occupations with the KldB-IDs ['63302', '63301', '63312', '63322', '63382']. The model has the following activity fields: ['H1: Reinigungsarbeiten durchführen (Housekeeping)', 'H2: Veranstaltungen vorbereiten und im Warenlager arbeiten', 'H3: In der Küche helfen', 'H4: Im Service arbeiten', 'H5: An der Rezeption arbeiten', 'H6: Service leiten']. For the goldstandard 748 documents were annotated by 2 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 H1: Reinigungsarbeiten durchführen (Housekeeping)	0.47	57
1 H2: Veranstaltungen vorbereiten und im Warenlager arbeiten	0.4	106
2 H3: In der Küche helfen	0.86	231
3 H4: Im Service arbeiten	0.75	464
4 H5: An der Rezeption arbeiten	0.93	30
5 H6: Service leiten	0.55	92

Tiefbaufacharbeiter/in

Slug: **tbf**

This model is for occupations with the KldB-IDs ['32222', '32201', '32202', '32232', '32252']. The model has the following activity fields: ['H1: Vermessungs- und Erdbauarbeiten durchführen', 'H2: Schächte setzen, Rohrleitungen verlegen und instand setzen', 'H3: Pflasterflächen herstellen', 'H4: Verkehrsflächen instand setzen', 'H5: Einfache Hochbauarbeiten durchführen']. For the goldstandard 1041 documents were annotated by 8 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 H1: Vermessungs- und Erdbauarbeiten durchführen	0.68	372
1 H2: Schächte setzen, Rohrleitungen verlegen und instand setzen	0.73	474
2 H3: Pflasterflächen herstellen	0.84	391
3 H4: Verkehrsflächen instand setzen	0.77	341
4 H5: Einfache Hochbauarbeiten durchführen	0.36	76

Hochbaufacharbeiter/in

Slug: **hbf**

This model is for occupations with the KldB-IDs ['32122', '32101', '32102', '32112', '32132', '32142', '32152', '32162', '32172']. The model has the following activity fields: ['H1: Einfache einschalige Baukörper erstellen', 'H2: Bauwerke erschließen', 'H3: Baukörper mit tragenden Funktionen herstellen', 'H4: Baukörper abdichten und dämmen', 'H5: Einfache Ausbauarbeiten durchführen', 'H6: Besondere Bauteile und Sichtmauerwerk herstellen']. For the goldstandard 965 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 H1: Einfache einschalige Baukörper erstellen	0.82	225
1 H2: Bauwerke erschließen	0.85	215
2 H3: Baukörper mit tragenden Funktionen herstellen	0.92	207
3 H4: Baukörper abdichten und dämmen	0.89	80
4 H5: Einfache Ausbauarbeiten durchführen	0.86	201
5 H6: Besondere Bauteile und Sichtmauerwerk herstellen	0.7	81

Ausbaufacharbeiter/in

Slug: abf

This model is for occupations with the KldB-IDs ['33112', '33101', '33102', '33122', '33132']. The model has the following activity fields: ['H1: Wandfliesen im Dünnbett verlegen', 'H2: Bodenfliesen im Dünnbett verlegen', 'H3: Zubehör und zusätzliche Materialien einbauen', 'H4: Spezielle Fliesen- und Plattenverlegearbeiten durchführen', 'H5: Dämm-, Trocken-, Putz- und Estricharbeiten ausführen', 'H6: Einfache Hochbauarbeiten durchführen']. For the goldstandard 894 documents were annotated by 4 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 H1: Wandfliesen im Dünnbett verlegen	0.7	290
1 H2: Bodenfliesen im Dünnbett verlegen	0.67	304
2 H3: Zubehör und zusätzliche Materialien einbauen	0.62	281
3 H4: Spezielle Fliesen- und Plattenverlegearbeiten durchführen	0.71	264
4 H5: Dämm-, Trocken-, Putz- und Estricharbeiten ausführen	0.8	193
5 H6: Einfache Hochbauarbeiten durchführen	0.55	26

Fachinformatiker/-in (3 Fachrichtungen – Systemintegration, Anwendungstechnik und digitale Vernetzung)

Slug: informatiker

This model is for occupations with the KldB-IDs ['43102', '43112', '43122', '43152', '43412']. The model has the following activity fields: ['HF1: Einfache IT-Systeme zusammenbauen, installieren und warten (A)', 'HF2: IT-Netzwerke installieren, integrieren und konfigurieren (B)', 'HF3: IT-Sicherheit anwenden, Serverdienste installieren und warten (C)', 'HF4: Mit Datenbanken arbeiten (D1, D2)', 'HF5: Schnittstellen programmieren (D3)', 'HF6: Anwendungen programmieren (E1, E2, F2)', 'HF7: Cyber-physischer-Systeme einrichten (E3)', 'HF8: Systeme automatisieren und pflegen (F1)', 'HF9: Cyber-physischer-Systeme optimieren (F3)']. For the goldstandard 481 documents were annotated by 2 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Einfache IT-Systeme zusammenbauen, installieren und warten (A)	0.864541	339
1 HF2: IT-Netzwerke installieren, integrieren und konfigurieren (B)	0.916679	322
2 HF3: IT-Sicherheit anwenden, Serverdienste installieren und warten (C)	0.898788	375
3 HF4: Mit Datenbanken arbeiten (D1, D2)	0.710678	146
4 HF5: Schnittstellen programmieren (D3)	0.653541	58
5 HF6: Anwendungen programmieren (E1, E2, F2)	0.595581	232
6 HF7: Cyber-physischer-Systeme einrichten (E3)	0.314394	12
7 HF8: Systeme automatisieren und pflegen (F1)	0.729331	87
8 HF9: Cyber-physischer-Systeme optimieren (F3)	0	2

Fachkraft Metalltechnik – Fachrichtung Konstruktionstechnik

Slug: metall_konstruktion

This model is for occupations with the KldB-IDs ['24411', '24412']. The model has the following activity fields: ['HF1: Bauteile mit handgeführten Werkzeugen und Maschinen herstellen', 'HF2: Einzelteile mit Werkzeugmaschinen drehen und fräsen', 'HF3: Baugruppen und Metallkonstruktionen montieren und demontieren', 'HF4: Bleche und Rohre umformen und trennen', 'HF5: Bauteile und Baugruppen schweißen']. For the goldstandard 490 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Bauteile mit handgeführten Werkzeugen und Maschinen herstellen	0.706	174
1 HF2: Einzelteile mit Werkzeugmaschinen drehen und fräsen	0.301	51
2 HF3: Baugruppen und Metallkonstruktionen montieren und demontieren	0.81	350
3 HF4: Bleche und Rohre umformen und trennen	0.687	158
4 HF5: Bauteile und Baugruppen schweißen	1	178

Fachkraft Metalltechnik – Fachrichtung Zerspanungstechnik

Slug: metall_zerspanung

This model is for occupations with the KldB-IDs ['24231', '24232']. The model has the following activity fields: ['HF1: Bauteile mit handgeführten Werkzeugen und Maschinen herstellen', 'HF2: Einzelteile mit Werkzeugmaschinen drehen und fräsen', 'HF3: Baugruppen und Metallkonstruktionen montieren und demontieren', 'HF4: CNC-Drehmaschinen und -Drehmaschinensysteme bedienen, überwachen und warten', 'HF5: CNC-Fräsmaschinen und -Fräsmaschinensysteme bedienen, überwachen und warten']. For the goldstandard 538 documents were annotated by 2 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Bauteile mit handgeführten Werkzeugen und Maschinen herstellen	0.682	63
1 HF2: Einzelteile mit Werkzeugmaschinen drehen und fräsen	0.553	407
2 HF3: Baugruppen und Metallkonstruktionen montieren und demontieren	0.717	14
3 HF4: CNC-Drehmaschinen und -Drehmaschinensysteme bedienen, überwachen und warten	0.605	496
4 HF5: CNC-Fräsmaschinen und -Fräsmaschinensysteme bedienen, überwachen und warten	0.354	480

Groß- und Außenhandelsmanagement

Slug: grosshandel

This model is for occupations with the KldB-IDs ['61212']. The model has the following activity fields: ['HF1: Waren beschaffen', 'HF2: Konten führen und Zahlungsverkehr abwickeln', 'HF3: Kunden beraten und Waren verkaufen', 'HF4: Im Großhandelslager arbeiten', 'HF5: Waren im Außenhandel exportieren']. For the goldstandard 582 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Waren beschaffen	0.57	342
1 HF2: Konten führen und Zahlungsverkehr abwickeln	0.828	308
2 HF3: Kunden beraten und Waren verkaufen	0.692	526
3 HF4: Im Großhandelslager arbeiten	0.633	396
4 HF5: Waren im Außenhandel exportieren	0.632	252

Industriemechaniker/in

Slug: **industriemechaniker**

This model is for occupations with the KldB-IDs ['25101', '25102', '25112']. The model has the following activity fields: ['HF1: Werkstücke durch manuelle und maschinelle Zerspanung herstellen und bearbeiten', 'HF2: Mechanische Verbindungen herstellen', 'HF3: Bauteile auf numerisch gesteuerten Werkzeugmaschinen herstellen', 'HF4: Steuerungen der Fluidtechnik aufbauen, prüfen und in Betrieb nehmen', 'HF5: Elektrotechnische Systeme installieren, prüfen und in Betrieb nehmen', 'HF6: Technische Systeme in Stand halten']. For the goldstandard 443 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Werkstücke durch manuelle und maschinelle Zerspanung herstellen und bearbeiten	0.63	111
1 HF2: Mechanische Verbindungen herstellen	0.66	311
2 HF3: Bauteile auf numerisch gesteuerten Werkzeugmaschinen herstellen	0.487	93
3 HF4: Steuerungen der Fluidtechnik aufbauen, prüfen und in Betrieb nehmen	0.801	110
4 HF5: Elektrotechnische Systeme installieren, prüfen und in Betrieb nehmen	0.315	43
5 HF6: Technische Systeme in Stand halten	0.763	281

Industrieelektriker/in

Slug: **industrieelektriker**

This model is for occupations with the KldB-IDs ['26252']. The model has the following activity fields: ['HF1: Komponenten und Betriebsmittel bearbeiten und montieren', 'HF2: Elektrische Anlagen und Maschinen installieren', 'HF3: Elektrische Anlagen, Maschinen und Geräte instand setzen', 'HF4: Elektrische Anlagen, Maschinen und Geräte in Betrieb nehmen']. For the goldstandard 557 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Komponenten und Betriebsmittel bearbeiten und montieren	0.77	252
1 HF2: Elektrische Anlagen und Maschinen installieren	0.775	279
2 HF3: Elektrische Anlagen, Maschinen und Geräte instand setzen	0.927	389
3 HF4: Elektrische Anlagen, Maschinen und Geräte in Betrieb nehmen	0.845	316

Fachlagerist/in

Slug: lager

This model is for occupations with the KldB-IDs ['51311', '51312']. The model has the following activity fields: ['HF1: Waren annehmen', 'HF2: Fördermittel einsetzen', 'HF3: Waren lagern', 'HF4: Waren kommissionieren', 'HF5: Waren versenden', 'HF6: Logistische Prozesse planen und optimieren', 'HF7: Waren beschaffen']. For the goldstandard 537 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Waren annehmen	0.866	312
1 HF2: Fördermittel einsetzen	0.839	297
2 HF3: Waren lagern	0.77	240
3 HF4: Waren kommissionieren	0.795	329
4 HF5: Waren versenden	0.775	398
5 HF6: Logistische Prozesse planen und optimieren	0.798	83
6 HF7: Waren beschaffen	0.494	11

Gärtner:in

Slug: gaertner

This model is for occupations with the KldB-IDs ['12101', '12102', '12142']. The model has the following activity fields: ['HF1: Erdarbeiten ausführen und Maschinen bedienen', 'HF2: Befestigte Flächen herstellen und Entwässerungssysteme einbauen', 'HF3: Bauwerke herstellen', 'HF4: Vegetationstechnische Arbeiten und Neupflanzungen durchführen', 'HF5: Gärtnerische Pflegemaßnahmen durchführen']. For the goldstandard 456 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Erdarbeiten ausführen und Maschinen bedienen	0.592	181
1 HF2: Befestigte Flächen herstellen und Entwässerungssysteme einbauen	0.753	150
2 HF3: Bauwerke herstellen	0.572	112
3 HF4: Vegetationstechnische Arbeiten und Neupflanzungen durchführen	0.752	156
4 HF5: Gärtnerische Pflegemaßnahmen durchführen	0.643	355

Tischler:in

Slug: tischler

This model is for occupations with the KldB-IDs ['22301', '22302', '22342']. The model has the following activity fields: ['HF1: Produkte an Holzbearbeitungsmaschinen herstellen', 'HF2: Einfache Produkte in Einzelfertigung herstellen und zusammenbauen', 'HF3: Oberflächen beschichten, lackieren und veredeln', 'HF4: Möbel demontieren und montieren', 'HF5: Bauelemente demontieren und montieren', 'HF6: Wartungs- und Instandhaltungsarbeiten durchführen']. For the goldstandard 262 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Produkte an Holzbearbeitungsmaschinen herstellen	0.749	159
1 HF2: Einfache Produkte in Einzelfertigung herstellen und zusammenbauen	0.412	127
2 HF3: Oberflächen beschichten, lackieren und veredeln	0.436	135
3 HF4: Möbel demontieren und montieren	0.644	144
4 HF5: Bauelemente demontieren und montieren	0.595	129
5 HF6: Wartungs- und Instandhaltungsarbeiten durchführen	0.709	59

Bäcker:in

Slug: **baecker**

This model is for occupations with the KldB-IDs ['29222']. The model has the following activity fields: ['HF1: Teige herstellen', 'HF2: Teige zu Brot und Kleingebäck aufarbeiten', 'HF3: Feine Backwaren aus Teigen aufarbeiten', 'HF4: Feine Backwaren aus Massen herstellen, Torten herstellen', 'HF5: Backwarensnacks und kleine Gerichte herstellen', 'HF6: Brot, Kleingebäck und Feine Backwaren abbacken']. For the goldstandard 391 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Teige herstellen	0.539	272
1 HF2: Teige zu Brot und Kleingebäck aufarbeiten	0.682	221
2 HF3: Feine Backwaren aus Teigen aufarbeiten	0.519	205
3 HF4: Feine Backwaren aus Massen herstellen, Torten herstellen	0.642	206
4 HF5: Backwarensnacks und kleine Gerichte herstellen	0.608	158
5 HF6: Brot, Kleingebäck und Feine Backwaren abbacken	0.631	261

Fachkraft Lebensmittelhandwerk (Bäckerei)

Slug: **fk_baeckerei**

This model is for occupations with the KldB-IDs ['62312']. The model has the following activity fields: ['HF1: Backwaren und Betriebsmittel annehmen, prüfen und lagern', 'HF2: Snacks, kleine Gerichte und Getränke herstellen', 'HF3: Teiglinge zu Backwaren abbacken', 'HF4: Kunden beraten, Waren verkaufen', 'HF5: Kassiovorgänge durchführen', 'HF6: Verkaufsfördernde Maßnahmen durchführen']. For the goldstandard 505 documents were annotated by 2 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Backwaren und Betriebsmittel annehmen, prüfen und lagern	0.917	300
1 HF2: Snacks, kleine Gerichte und Getränke herstellen	0.944	248
2 HF3: Teiglinge zu Backwaren abbacken	0.78	176
3 HF4: Kunden beraten, Waren verkaufen	0	439
4 HF5: Kassiovorgänge durchführen	0.963	222
5 HF6: Verkaufsfördernde Maßnahmen durchführen	0.571	286

Maler:in

Slug: maler

This model is for occupations with the KldB-IDs ['33211', '33212']. The model has the following activity fields: ['HF1: Innen und außen anstreichen', 'HF2: Objekte lackieren', 'HF3: Putzarbeiten durchführen', 'HF4: Trockenbauarbeiten durchführen', 'HF5: Tapeten und Bodenbeläge kleben', 'HF6: Dämmarbeiten durchführen']. For the goldstandard 518 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Innen und außen anstreichen	0.64	410
1 HF2: Objekte lackieren	0.599	391
2 HF3: Putzarbeiten durchführen	0.813	218
3 HF4: Trockenbauarbeiten durchführen	0.765	298
4 HF5: Tapeten und Bodenbeläge kleben	0.794	341
5 HF6: Dämmarbeiten durchführen	0.974	100

Gebäudereiniger:in

Slug: reiniger

This model is for occupations with the KldB-IDs ['54101', '54112', '54132', '54142', '54152', '54182']. The model has the following activity fields: ['HF1: Unterhaltsreinigung durchführen', 'HF2: Glasflächen, Fenster und Rahmen reinigen', 'HF3: Desinfizierende Reinigung durchführen ', 'HF4: Fassaden- und Industriereinigung durchführen', 'HF5: Sonderreinigungen durchführen']. For the goldstandard 521 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Unterhaltsreinigung durchführen	0.526	427
1 HF2: Glasflächen, Fenster und Rahmen reinigen	1	50
2 HF3: Desinfizierende Reinigung durchführen	0.582	21
3 HF4: Fassaden- und Industriereinigung durchführen	0.458	61
4 HF5: Sonderreinigungen durchführen	0.596	56

Berufskraftfahrer:in

Slug: kraftfahrer

This model is for occupations with the KldB-IDs ['52122', '52132']. The model has the following activity fields: ['HF1: Touren planen und dokumentieren', 'HF2: Fahrzeug vorbereiten', 'HF3: Güter befördern', 'HF4: Fahrzeug führen', 'HF5: Fahrzeug warten und pflegen', 'HF6: Personen befördern']. For the goldstandard 589 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Touren planen und dokumentieren	0.569	121
1 HF2: Fahrzeug vorbereiten	0.752	409
2 HF3: Güter befördern	0.731	441
3 HF4: Fahrzeug führen	0.495	575
4 HF5: Fahrzeug warten und pflegen	0.864	148
5 HF6: Personen befördern	0.884	71

Kaufmann/Kauffrau für Spedition und Logistikdienstleistung

Slug: spedition

This model is for occupations with the KldB-IDs ['51622']. The model has the following activity fields: ['HF1: Logistische Leistungen anbieten', 'HF2: Frachtaufträge im Güterkraftverkehr disponieren', 'HF3: Speditionsaufträge im Sammelgut- und Systemverkehr bearbeiten', 'HF4: Luft- und Seefrachtaufträge bearbeiten', 'HF5: Außenhandelsaufträge abwickeln', 'HF6: Geschäftsvorgänge steuern und kontrollieren']. For the goldstandard 569 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Logistische Leistungen anbieten	0.324	500
1 HF2: Frachtaufträge im Güterkraftverkehr disponieren	0.508	356
2 HF3: Speditionsaufträge im Sammelgut- und Systemverkehr bearbeiten	0.262	360
3 HF4: Luft- und Seefrachtaufträge bearbeiten	0.67	255
4 HF5: Außenhandelsaufträge abwickeln	0.817	333
5 HF6: Geschäftsvorgänge steuern und kontrollieren	0.776	293

Fachkraft für Schutz und Sicherheit

Slug: sicherheit

This model is for occupations with the KldB-IDs ['53111', '53112']. The model has the following activity fields: ['HF1: Menschen und Objekte schützen', 'HF2: Schutz- und Sicherungsmaßnahmen im öffentlichen (Verkehrs)Raum durchführen', 'HF3: Veranstaltungsdienst durchführen', 'HF4: Sicherheitstechnische Einrichtungen und Hilfsmittel einsetzen und anwenden', 'HF5: Sicherheitsrelevante Sachverhalte ermitteln, aufklären und dokumentieren', 'HF6: Sicherheitskonzepte erstellen und Geschäftsprozesse durchführen']. For the goldstandard 486 documents were annotated by 3 annotators. 20% of documents were annotated by at least two annotators to calculate the inter-annotators-agreement.

activity_field	Krippendorff's Alpha	Occurrences in Goldstandard
0 HF1: Menschen und Objekte schützen	0.594	405
1 HF2: Schutz- und Sicherungsmaßnahmen im öffentlichen (Verkehrs)Raum durchführen	0.776	47
2 HF3: Veranstaltungsdienst durchführen	0.843	35
3 HF4: Sicherheitstechnische Einrichtungen und Hilfsmittel einsetzen und anwenden	0.73	161
4 HF5: Sicherheitsrelevante Sachverhalte ermitteln, aufklären und dokumentieren	0.493	14
5 HF6: Sicherheitskonzepte erstellen und Geschäftsprozesse durchführen	0.683	35

Dataset Creation

Curation Rationale

The data was collected from the “Jobbörsen” of the “Bundesagentur für Arbeit”. Before annotation rounds, all job ads from the last seven days for the given occupation were collected. As input, the matching KldB-8-IDs for their respective KLDB-5-IDs were supplied for the search. The list of matching occupations was curated by an industry expert.

Source Data

The job ads are written by the employing organizations. Using the Jobbörsen API, the data quality is high and minimal mismatches between occupations and job ads were found (These were excluded from the gold standard).

Annotations

Annotation process

Each occupation model was annotated separately and followed roughly the same process: 1. An industry expert introduces the activity field model to the annotators, and the first annotation rules are written down. 2. In the first round, 50 documents are annotated by all annotators (For most rounds, about 3 annotators). 3. In an adjudication meeting, the annotation quality is discussed using Fleiss' Kappa and inter-annotator agreement metrics. The annotation guidelines are adapted as necessary. 4. The corpus of documents is assigned to the annotators with an overlap of 10% to 20% to assess the annotation quality for the gold standard. 5. The overlapping documents are adjudicated using a merge strategy (this results in a slight decrease in overall annotation quality; a full manual adjudication, however, was not possible due to time constraints).

Who are the annotators?

Annotators are all employees of the Bertelsmann Foundation and were trained for the annotation task by industry experts.

Personal and Sensitive Information

The OJAs partially contain the address and contact information of the person responsible for HR in the respective organization. This information was not used in model training and cannot be extracted from the final model.

Considerations for Using the Data

Generally, a model trained on this gold standard should only be used for German online job ads. **Do not extract activity fields of an OJA for an occupation it was not intended for by the publisher of the OJA.** This will result in meaningless predictions!

Social Impact of Dataset

The present data set is the only one that features annotations of activity fields in online job ads. It is a valuable contribution to do more research on the demand for activity fields. It can also help improve labor market policies and support educational institutions.

Discussion of Biases

The data was annotated by over ten different annotators. While the annotators themselves are not experts or practitioners in the occupations, they were briefed and continuously supported by experts who developed the activity field model in Germany.

Other Known Limitations

The dataset features only German online job ads. However, there are also many job ads that are published in English, which cannot be analyzed using this data set. Some activity fields are rare in practice and, therefore, rare in online job ads. Since fewer annotations exist for these concepts, any model has to face the challenge of unbalanced data.

Additional Information

Dataset Curators

The dataset was curated by Johannes Müller (johannes.mueller@and-effect.com).