

# Classification of German job titles in online job postings using the KldB- 2010 taxonomy

Technical Report

**Version**

1

**Authors**

Rahkakavee Baskaran, Johannes Müller

**Created at**

June 28, 2023

**Last updated at**

November, 7, 2023

The logo for 'effect' features a stylized ampersand symbol followed by the word 'effect' in a bold, lowercase, sans-serif font.

# Table of Contents

<b>1</b>	<b>Summary .....</b>	<b>5</b>
<b>2</b>	<b>Methods.....</b>	<b>5</b>
2.1	Taxonomy .....	5
2.2	Model requirements .....	9
2.3	Data Basis, Challenges and Limitations.....	10
2.4	Data Cleaning Process and Taxonomy Processing.....	12
2.5	Extraction pipeline .....	15
2.6	Filter of not classifiable job titles .....	15
2.7	Semantic Search .....	16
<b>3</b>	<b>Model .....</b>	<b>17</b>
3.1	Model Details .....	17
3.2	Intended Use .....	18
3.3	Metrics.....	18
3.4	Training Data .....	19
3.5	Evaluation Data .....	20
3.6	Results Fine Tuning Process .....	20
3.7	Results Classification Task .....	21
<b>4</b>	<b>Application .....</b>	<b>22</b>

# List of Figures

Figure 1: Input and Output of Algorithm	9
Figure 2: Semantic Search	17
Figure 3: Fine-Tuning Evaluation Results	21
Figure 4: Top 10 KLDB Predictions	23
Figure 5: Distribution of KLDBs of Top 50 Jobs assigned with mapping	24
Figure 6: Distribution of KldBs of Top 50 Jobs assigned with algorithm	24
Figure 7: Distribution of titles mapped to "Haus-und Familienpflege-Fachkraft"	25

# List of Tables

Table 1: Overview of the KldB Level 1 names	6
Table 2: Overview of the KldB Levels 2010 (as of June 2022)	7
Table 3: Number of categories per KldB Level after Filtering (as of June 2022)	9
Table 4: Evaluation of Fine Tuning	20
Table 5: Comparison of Mapping and Algorithm Predictions	26
Table 6 (Appendix): General Evaluation Rules Data Cleaning	27
Table 7 (Appendix): Specific Evaluation Rules Data Cleaning	29
Table 8(Appendix): Rules Sachbearbeiter:in	40
Table 9 (Appendix): Rules Ingenieur:in	43
Table 10 (Appendix): Rules Referent:in	44
Table 11 (Appendix): Rules Assistent:in	45
Table 12 (Appendix): Rules Consultant:in	45
Table 13 (Appendix): Baseline Evaluation Results	47

# List of abbreviations

BA	Bundesagentur für Arbeit
KIdB	Klassifikation der Berufe 2010 – überarbeitete Fassung 2020
BKZ	Berufskennzahl
ISCO	International Standard Classification of Occupations
ID	Identification
OJA	Online Job Advertisement
O*NET	Occupational Information Network

# 1 Summary

The occupation classification algorithm is used to classify non-normalized occupational titles from online job advertisements (OJA) according to the "Klassifikation der Berufe 2010 – überarbeitete Fassung 2020" (KldB)<sup>1</sup>. The algorithm is based on a Bidirectional Encoder Representations from Transformers (BERT) model. The classification is performed using semantic search. The data is evaluated using data from the "Jobbörse" of the "Bundesagentur für Arbeit"<sup>2</sup> (BA). To evaluate and develop the algorithm, the data had to be cleaned and partially recoded. The classification has an accuracy of 0.86 and a macro f1 score of 0.70 on the five digits of the KldB classification.

## 2 Methods

### 2.1 Taxonomy

The occupational classification of the OJA's is based on the individual occupations in the so-called BKZ ("Berufskennzahl"<sup>3</sup>) and their grouping in the so-called KldB 2010 ("Klassifikation der Berufe 2010"<sup>4</sup>) developed by the BA. Each occupation consists of an eight-digit ID and a long and short label name. In the following, the eight-digit ID is referred to as the BKZ ID, and the long label name of the individual occupation is the BKZ name. The short name will not be considered further. The complete collection of individual occupations is referred to as the BKZ taxonomy. The first five digits of the BKZ ID are used to group the individual occupations hierarchically in various dimensions into larger occupational groups. This systematic grouping of individual occupations is described in the KldB 2010 taxonomy. As for the individual occupations, there is a long and a short label name for each group. Subsequently, the first five digits of the BKZ ID are referred to as the KldB ID. The respective short-label names are denoted as KldB names. In the following, first the systematic behind the KldB taxonomy is described and explained. Thereafter, follows a brief description of the BKZ taxonomy.

---

<sup>1</sup> Classification of Occupations 2010 – revised version 2020

<sup>2</sup> Eng. Federal Employment Agency

<sup>3</sup> Eng. occupation code

<sup>4</sup> Eng. Classification of professions 2010

Systematics of KldB. The KldB taxonomy represents a classification scheme for Germany, reflecting the current trends in the labor market, and is based on a sophisticated, statistical procedure. Containing valuable information, like the vertical requirement dimensions, the taxonomy is well-suited for analysis and further downstream tasks. Furthermore, the KldB taxonomy is linked to the International Standard Classification of Occupations (ISCO), which enables analysis and comparisons on an international level. (Baskaran, 2022)

The KldB taxonomy is hierarchically divided into five levels. At each level, the individual occupations are summarized, with the levels becoming increasingly detailed from the highest to the lowest. The labels of the KldB are coded using a five-digit number. The top hierarchy level, “Berufsbereich” (Level 1), divides occupations into ten broader categories. The level one KldB ID has a length of one and ranges from zero to nine. Table 1 shows a list of ten KldB IDs for level one and their names.

Table 1: Overview of the KldB Level 1 names

KldB 1 <sup>st</sup>	KldB name
Level	
0	Militär
1	Land-, Forst- und Tierwirtschaft
2	Rohstoffgewinnung, Produktion und Fertigung
3	Bau, Architektur, Vermessung und Gebäudetechnik
4	Naturwissenschaft, Geografie und Informatik
5	Verkehr, Logistik, Schutz und Sicherheit
6	Kaufmännische Dienstleistungen, Warenhandel, Vertrieb, Hotel und Tourismus
7	Unternehmensorganisation, Buchhaltung, Recht und Verwaltung
8	Gesundheit, Soziales, Lehre und Erziehung
9	Sprach-, Literatur-, Geistes-, Gesellschafts- und Wirtschaftswissenschaften, Medien, Kunst, Kultur und Gestaltung

Each of the level-one classes is subdivided into one or more subclasses, thereby creating level two of the KLDB taxonomy. Level two has an ID length of two, where the first digit denotes level one, and the second digit denotes level two. For example, the “Berufsbereich” “Verkehr, Logistik, Schutz und Sicherheit”, with KldB ID 5 on level one, has four subclasses. The first subclass is “Verkehrs- und Logistikberufe (außer Fahrzeugführung)” and has the KldB ID 51. Similarly, level three further subdivides the classes of level two, and this process finally results in a five-digit KldB ID at level five. We can tell from the KldB ID 54193 (KldB name: Aufsichtskräfte - Reinigung), for example, that the level one KldB ID is 5, and the second level KldB ID is 4. An occupation or job title can be assigned to all five levels. Consider the job title 'Helfer/in - Lagerwirtschaft, Transport.' The job title should be assigned to the KldB ID 51311, which stands for "Lagerwirtschaft - Helfer." Using the five-digit KldB ID, it is possible to assign the title, for instance, to the Level one KldB ID 5 (Berufsbereich "Verkehr, Logistik, Schutz und Sicherheit") and the Level three KldB ID 513 ("Lagerwirt., Post, Zustellung, Güterumschlag"). A detailed overview of each level, the number of classes, and an example are given in Table 2. Note that according to Bundesagentur für Arbeit (2021), the BA updates from time to time the occupations such, that some KldB levels become inactive. The BA provides a file with all labels BKZ labels, which are introduced in the next section with information on the current state of the occupations in their download portal (Bundesagentur für Arbeit, 2023a). We are using a version downloaded in June 2022, and since the KLDB was revised in 2021, there should be no major changes to the file.

Table 2: Overview of the KldB Levels 2010 (as of June 2022)

KldB Level	Level name	Number of categories	Example: “Helfer/in Lagerwirtschaft, Transport”
1	Berufsbereich	10	5: Verkehr, Logistik, Schutz und Sicherheit
2	Berufshauptgruppe	37	51: Verkehrs- und Logistikberufe (außer Fahrzeugführung)
3	Berufsgruppe	144	513: Lagerwirt., Post, Zustellung, Güterumschlag
4	Berufsuntergruppe	702	5131: Berufe in der Lagerwirtschaft
5	Berufsgattung	1300	51311: Lagerwirtschaft - Helfer

Further, the KldB taxonomy represents two dimensions: the horizontal dimension “Berufsfachlichkeit” and the vertical dimension “Anforderungsniveau”. Levels two to four are defined by “Berufsfachlichkeit.” The term “Berufsfachlichkeit” refers to the matching of “capabilities, skills, and knowledge” (dt. Fähigkeiten, Fertigkeiten und Kenntnisse (Paulus & Matthes, 2013, S. 8) between two occupations.<sup>5</sup>

The vertical dimension “Anforderungsniveau” (requirement level) structures the last level of the KLDB ID. The fifth digit of the KldB id is coded from zero to four, representing the complexity of the occupation. The first requirement level, “Helfer -und Anlern Tätigkeiten” represents an occupation with no specific educational training (dt. “berufliche Ausbildung”). “Fachlich ausgerichtete Tätigkeiten”, the second requirement level, defines occupations with at least two years of educational training (dt. Zweijährige Ausbildung). Requirement levels three and four define more complex occupations, with level three requiring a higher education training (dt. Meister- oder Technikerausbildung) or university degree and level four requiring a university degree with at least four years of studying. The vertical level can be used to group the occupations according to their requirement level and opens possibilities for various analyses (Paulus & Matthes, 2013; Baskaran, 2022; Bundesagentur für Arbeit, 2021).

Systematics of BKZ. The BKZ taxonomy represents the individual occupations. In contrast to the KLDB names, the BKZ names have more precise names, which are semantically closer to the actual occupations. For example, the occupation “Data Scientist” is assigned in the KLDB taxonomy to “Informatik (o.S.) - Experte” (43104), while the BKZ name is “Data Scientist” (43104132). Note that the first five digits of the BKZ ID are identical to the KLDB ID. Thus, it is also possible to infer from the BKZ ID to the KLDB ID. More than one BKZ ID can belong to one KLDB ID at a time. The 6<sup>th</sup> digit of the BKZ ID is used for differentiating between “Tätigkeiten” (job activities), coded with 1 and 2, and “Ausbildung”<sup>6</sup>, coded with 8 and 9. The last two digits of the BKZ ID are randomly assigned to differentiate the individual occupations. Further, each BKZ is assigned to two states: “E” and “R”. “E” stands for current BKZ and “R” for inactive BKZ. The states are not represented in the BKZ ID coding (Paulus & Matthes, 2013; Vicari, 2014). Lastly, each BKZ is assigned to one “berufskundliche Gruppe”<sup>7</sup>. The “berufskundliche Gruppe” classifies occupations according to their access requirements (required qualification) in different groups. Group A, for example, defines “Allgemeinbildung” and B defines “spezielle Berufskunde (Tätigkeiten)” (Vicari, 2014). In total, the BKZ Taxonomy contains 31245 BKZ labels (as of June 2022).

---

<sup>5</sup> In order to group the individual occupations into level 2-4 with the Berufsfachlichkeit a Cluster Analysis based on a table of competencies („Kompetenztabelle“, (Paulus & Matthes, 2013, S. 8)) of the BA was used.

<sup>6</sup> Eng. vocational education

<sup>7</sup> Eng. professional group



Processing Taxonomy. For creating the taxonomy for the algorithm, we used only labels from the BKZ taxonomy which are actively used (Status: E), and labels belonging to the “berufskundliche Gruppe” B. This results in 4.291 BKZ labels. This also reduces the number of labels for the KLDB Taxonomy as stated in Table 2. The updated numbers are provided in the following Table:

Table 3: Number of categories per KldB Level after Filtering (as of June 2022)

KldB Level	Level name	Number of categories
1	Berufsbereich	10
2	Berufshauptgruppe	37
3	Berufsgruppe	144
4	Berufsuntergruppe	691
5	Berufsgattung	1211

## 2.2 Model requirements

The goal of the occupation extractor algorithm is to classify extracted but not normalized job titles of OJA’s according to the KLDB ID on level 5.

The input and output of the algorithm are defined as in the following example:

```
input = [
{
  "posting_id": "7e94c942-39a9-b139-2c60-8b685e87382d", "job_title": "Juniorverkäufer BMW (m/w)"
},
{
  "posting_id": "50488e78-2670-5346-36d6-2a62017d293c", "job_title": "Anwendungstechniker/Application Engineer (m/w) Robotik"
},
{
  "posting_id": "7f7b4786-c190-4dc4-07db-45786b69ee76", "job_title": "Vertriebsassistent (m/w)"
}
]

output = [
{
  "posting_id": '7e94c942-39a9-b139-2c60-8b685e87382d', 'kldb_id': '62272'
},
{
  "posting_id": '50488e78-2670-5346-36d6-2a62017d293c', 'kldb_id': '43412'
},
{
  "posting_id": '7f7b4786-c190-4dc4-07db-45786b69ee76', 'kldb_id': '61122'
}
]
```

Figure 1: Input and Output of Algorithm

The overall performance goal of the algorithm is the correct identification of all five digits of the KLDB – with positions one, three, and five being most important for analysis. Although the first level is not integrated into the two dimensions of the KLDB taxonomy, it is especially relevant because the correctness of the first level determines whether the classification of the job title can be correct at the other levels. Accordingly, high correctness at the first level is crucial for the high performance of the algorithm. The third and fifth levels of the KLDB are considered particularly relevant for analytical tasks due to their informative content in their grouping.

We define the following performance for each of the three digits:

- KLDB Level one: at least 85% for all metrics
- KLDB Level three: at least over 75% for all metrics, except the accuracy should be at least 85%
- KLDB Level five: at least over 70% for all metrics, except the accuracy should be at least 85%

Further, there are some, qualitative extraction quality requirements for the classification, which must be checked manually:

Firstly, the algorithm must be able to identify the correct label for job titles explicitly containing one of the BKZ names or which are semantically close to one of the BKZ names. At least the BKZ name should be in the correct domain. An example from the baseline approach showed that for the title “Neurologe/in” the baseline predicted “Japanaloge/in” instead of “Facharzt/-ärztin – Neurologie”.

Secondly, it is important that the requirement level (the fifth digit of the KldB ID) is predicted correctly if the job title includes words that explicitly refer to one of the requirement levels. For example, a job title containing the word “Helfer:in” must be assigned to a KldB ID with the last digit being a one.

Lastly, some professions are in greater demand than others. It must be ensured that large occupational groups are correctly classified to avoid strong biases due to false predictions.

## **2.3 Data Basis, Challenges and Limitations**

The data basis for training the algorithm is scraped from the “Jobbörse” (Bundesagentur für Arbeit, 2023b) of the BA. The BA is one of the biggest job portals on the German job market. Compared to other portals, the BA uses the KldB taxonomy to structure the vacancies. Each vacancy has one of the BKZ ids and the job title, which form the basis of our training dataset. As described in Chapter 2.1 the KldB id can be inferred from the BKZ ID.

In total, we scraped vacancies for a period between 2020 and 2023 on an irregular basis. Combining all these datasets, the original dataset contains 62.320 entries. Initial attempts to use this complete dataset and train different deep learning models have led to a performance of around 60% accuracy but lower precision, recall, and f1 macro scores. Although we used different techniques for building the algorithm, it was not possible to improve the performance with this dataset, which suggests poor data quality. In fact, manual checks revealed inconsistencies and wrong assignments. Furthermore, the data is highly imbalanced, which leads to low coverage of the taxonomy labels. In the subsequent section, a comprehensive examination of each matter, accompanied by exemplifications, is presented.

Inconsistency and wrong assignments. The assignment of the bkz label to the job title is provided by the employer or a service agency. Given the number of labels in the bkz taxonomy, this could lead to various inconsistencies. Firstly, similar job titles or titles with the same occupation can be labeled differently by employers. The original dataset contains, for instance, twice the job title “Baumkontrolleur (d/m/w)”. Both are labeled with different bkz names: “Baumpfleger/in” and “Forstwirt/in”. These inconsistencies and ambiguities must not be errors per se as the same job title could be used for different occupations.<sup>8</sup> However, to create a labeling system that only relies on the job title, these ambiguities influence the evaluation substantially.

Secondly, as described in Chapter 2.1, the fifth digit of the BKZ ID describes the requirement level. The original dataset is inconsistent regarding this requirement level. For instance, titles that contain the word “Helfer:in” must be assigned to a bkz, with the fifth digit being 1. Similarly, titles containing “Fachkraft” cannot have a 1 as the fifth digit. Same as for the fifth digit, the fourth digit of the BKZ ID indicates with the digit 9 if the occupation is a leader position. If a job title contains the word “Leitung”, then it cannot be assigned to a bkz id without a 9 as the fourth digit.

Third, some job titles, for example, “Helfer:in”, “Mitarbeiter:in” or similar titles, cannot be assigned to a label in the taxonomy. However, each entry in the record is labeled with a BKZ label, leading to further inconsistent and incorrect data. Lastly, there are plenty of entries with wrong assignments. For example, the job title “Bilanzbuchhalter Finanzbuchhalter m/w/d Gesundheitswesen” in the original dataset is assigned to the bkz label “Fachapotheker/in”.

---

<sup>8</sup> For an analysis of the ambiguity of job titles see Baskaran 2022.

Imbalanced data. The dataset is highly imbalanced. Out of the 4.291 labels in the taxonomy, after cleaning the data regarding the inconsistency, 3.192 labels were still missing in the dataset. Even though our machine learning approach does not require each label to be represented in the training data, it lowers performance, which is why it is necessary to have high coverage of the taxonomy in the training data. It is, however, difficult to augment the data for missing labels. Since targeted scraping for each label is not possible, augmenting all labels is a lot of manual work and not efficient. Besides this, for some labels, it is impossible to find real-world examples. This is because some BKZ labels are too detailed, or jobs are generally not advertised under this name. For example, for the occupation “Tierwirtschaftsmeister/in” the taxonomy specifies for the different animals detailed labels like “Tierwirtschaftsmeister/in – Rinderhaltung” or “Tierwirtschaftsmeister/in – Geflügehaltung”. It is unlikely that vacancies will list jobs with such detailed information in the job title or even in the job posting itself.

## 2.4 Data Cleaning Process and Taxonomy Processing

We utilized various steps, including several rounds of manual annotations, to address the above-described problems of the dataset and generate a gold standard training dataset for the training.

Remove inconsistencies and wrong assignments. In the first step, we manually scanned the original dataset and removed anomalies that we noticed at first glance. In the second step, we randomly sampled 18.231 entries from the dataset. Two annotators evaluated this dataset for correctness. Note that the data was evaluated in this step only, and no correction of the data was performed. For the evaluation, the dataset was first split into subsets according to the first digit of the bkz id and sorted by the bkz id. The large number of labels makes it impossible to remember all the labels when evaluating the dataset. By splitting and sorting the data, the process of identifying inconsistencies was much easier. Further, one subset, thus one occupation area, was only annotated by one annotator. This allowed the annotators to focus on one part of the taxonomy, reducing the risk of introducing new inconsistencies in the data. For the evaluation process, we used five labels:

- ☐ Label “0”: assignment of the original dataset is entirely wrong.
- ☐ Label “1”: the assignment of the original dataset is correct.
- ☐ Label “2”: the title in the original dataset cannot be assigned because there is no general name for that occupation in the taxonomy, for example: “Glaser:in”, which is not in the taxonomy, but there are several more specified “Glaser:in” in the taxonomy like “Glaser/in – Fenster- und Glasfassadenbau”.
- ☐ Label “-999”: the title in the original dataset cannot be assigned.

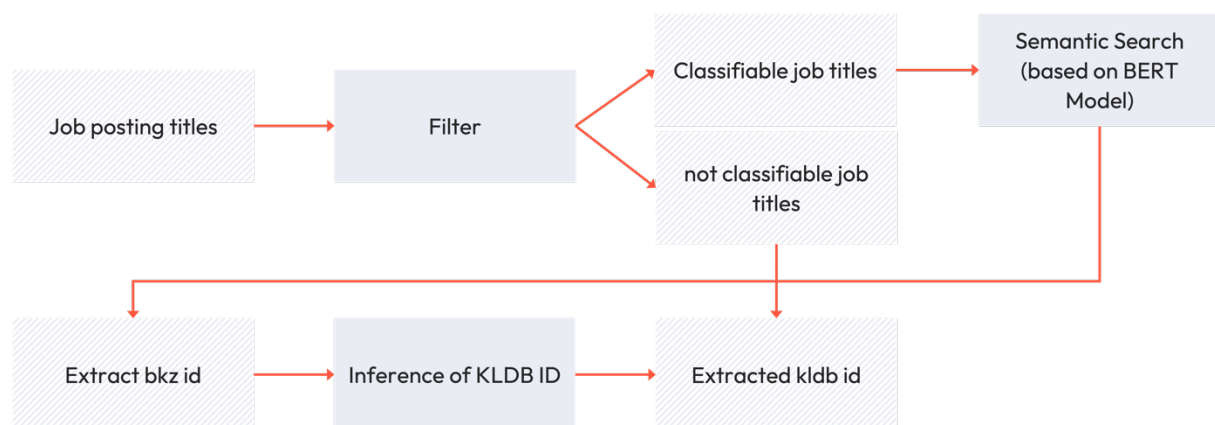
- Label “recode”: systematic inconsistent titles. Some titles containing specific words like “Sachbearbeiter” or “Ingenieur” were systematically inconsistent. In the evaluation process, we identified these and labeled them all with “recode” to manually correct them later. The following titles are affected by this: “Ingenieur”, “Sachbearbeiter”, “Consultant”, “Referent”, “Berater”, “Pflegekraft”, “BKZ Eisenbahn”, “Beamte”, “Elektroniker/Elektriker”, “BTA, CTA, PTA”, “Biologe”.

The subsets were divided between the annotators. During the evaluation, annotation rules for each subset and general annotation rules were established. For example, a general rule is “If a bkz label is word for word in the job title, then all other bkz labels are wrong”. A complete overview of all used annotation rules can be found in the appendix. After each subset, the annotation rules were discussed again, and the subsets were revised if necessary. In the third step, we merged all subsets and first excluded all entries which were evaluated with Label 0, -999, and 2. This affects 8.370 entries. From 18.231, 7.675 entries were labelled as correct (Label 1). Those entries were adopted directly into the gold standard training dataset. As a last step, entries with the label recode (2186 entries) were manually corrected by one annotator. The labels were corrected with the expert knowledge of the annotator. Furthermore, the annotator used the portal search “Jobbörse” to select the correct label. The portal search allows for searching with bkz labels or by typing in some keywords or titles. This helped to get an orientation of the frequency of assignment of titles to certain bkz labels. As for the first manual evaluation, a set of annotation rules was generated during the recoding process to ensure consistency. For example, all job titles like “Sachbearbeiter (m/w/d) Buchhaltung” were recoded to the bkz label “Buchhalter/in”. **A complete list of all the rules is included in the appendix.** Adding the corrected label, the dataset increased to a size of 9.500.

Address the imbalance in the data. As described in Chapter 2.2 the goal of the algorithm is to predict the KldB ID with five digits. However, we use the BKZ name for the prediction since it is semantically closer to the job titles. While for the KldB ID, approximately half of the labels are not represented in the taxonomy, for the BKZ ID only a quarter of the labels are covered in the cleaned dataset. With alignment to the goal, it is not relevant that all bkz labels are in the taxonomy if all KldB IDs are represented. We, therefore, excluded some bkz labels in such a way that each KldB ID is represented at least by one BKZ ID in the taxonomy. There are two reasons why we excluded the labels: first, as described above, data augmentation for all labels is not possible. By excluding some labels, we can efficiently do augmentation for those labels where the augmentation would improve the performance. Second, reducing the number of labels for the semantic search algorithm helps to increase performance since there is a much lower number of labels. We proceeded as follows to create the reduced taxonomy: first, we adopted all bkz labels, which are represented in the cleaned dataset, directly into the corpus. In a second step, we summarized all bkz labels that are missing in the cleaned dataset according to their KldB ID. For example, of all bkz labels belonging to the KldB label 11302, the bkz labels 'Pferdewirt/in (mit Schwerpunkten)', 'Pferdewirt/in – Pferdehaltung' und Service' are not represented in the cleaned dataset. After the summarization, we have 1.065 KldB labels containing BKZ labels that are not represented in the cleaned dataset. Out of these 1065 labels, 402 KldB labels had only one BKZ label. Those were also directly adopted into the reduced version of the corpus. In the next step, we went through the remaining KldB labels and, based on various criteria, excluded one or more BKZ labels or combined the labels into a more general label. For example, we excluded the label "Pferdewirt/in (mit Schwerpunkten)" from KldB 11302. For the KldB 11293, which contains the missing BKZ labels 'Tierwirtschaftsmeister/in – Rinderhaltung', 'Tierwirtschaftsmeister/in – Schäferei', 'Tierwirtschaftsmeister/in – Geflügelhaltung', 'Tierwirtschaftsmeister/in – Pelztierhaltung', 'Tierwirtschaftsmeister/in – Schweinehaltung', 'Tierwirtschaftsmeister/in – Imkerei', we built a new general label "Tierwirtschaftsmeister/in". To trace to which KldB the new label belongs, which is important for the prediction of the algorithm, we have taken the fifth digit number of the KldB and appended three zeros. In this case, the new label "Tierwirtschaftsmeister/in" receives the BKZ ID 11293000. In some cases, our criteria lead to excluding all labels. To ensure that all KldB ids were still represented in the taxonomy, we checked for those labels if there was another BKZ label for the same KldB label which is already in the cleaned dataset. If not, we kept all BKZ labels of this KldB ID. After the cleaning, the taxonomy could be reduced from an initial size of, 4291 to a taxonomy with 2690 BKZ labels without removing any KldB IDs. Although those labels are excluded and not used for the prediction, we make sure in the training process that the excluded labels are introduced to the model, which is described in Chapter 3.4 in more detail.

## 2.5 Extraction pipeline

The extraction pipeline is composed of three steps. First, the input, which is a list of job titles as described in Chapter 2.2, is passed through a filter. The filter takes each input title and compares it, based on simple matching, with a list of unclassifiable job titles. If there is no match for an input title, then this input title is passed to the semantic search model. Job titles that are not classifiable are directly passed to the output list. The Kldb ID is set to None for those titles. In the second step, a trained BERT model predicts with semantic search a BKZ ID based on the input title. In the last step, with the help of the BKZ ID the Kldb ID of the job title is inferred and builds the output of the pipeline. The diagram below gives an overview of this process:



## 2.6 Filter of not classifiable job titles

As discussed in Chapter 2.3, some titles cannot be classified. The job title “Gesucht Helfer(m/w/d)”, for example, cannot be assigned to one of the BKZ names since there is no general occupation “Helfer/in” listed in the BKZ taxonomy. The semantic search model outputs, due to its functionality, always one bkz name. Therefore, we use a filter before the semantic model to separate classifiable and non-classifiable titles. The classifiable titles are passed to the BERT model. Titles, which are not classifiable, are directly added to the output data with None as the Kldb ID. The filter is based on a list with unclassifiable titles like “Helfer/in” and a simple match of whether the input job title is in the list or not. This means we cannot filter out all unclassifiable cases. For example, the title “13 € Std. 4 Fleißige Bienchen (m/w/d) gesucht ab sofort” cannot be excluded in that way. However, the filter is also only meant to filter out job titles that occur repeatedly, such as “n job/s” or “Helfer/in”.

To create the list, we used different data sources. First, we sampled data from Textkernel Data (Textkernel, 2022). Textkernel scrapes OJA data from various sources. The data comes with different meta-data, like the publication date, the source of the OJA, and a not-normalized job title. In addition, the Textkernel data contains a mapping of the O\*NET (Occupational Information Network) professions to KldB taxonomy (O\*Net OnLine 2023). We draw two samples: One sample was randomly drawn and contains 10.000 examples. The mapping of O\*NET and KldB is such that not all titles in the dataset have a KldB ID. We thoroughly analyzed these titles and extracted candidate titles, from which we can later generate a list of not-classifiable job titles. Job titles with a KldB ID were only scanned. In addition, we draw a second sample with job titles that have no assignment according to O\*NET. In the same way as in the other example, we extracted possible candidates. Besides these two samples, we also extracted candidates from the list of job titles that were evaluated with -999 in the second step of the data cleaning process, described in 2.4. Next, we derived from the candidate list a list of words like “Helfer”, “Mitarbeiter” etc. Based on this list we generated job titles with different spellings like “Helfer/in”, “Helfer:in”, “Helfer/in gesucht” or “Helfer/in (m/w/d)”. This list is then used for filtering.

## 2.7 Semantic Search

The classification is realized using semantic search (SBERT, 2023) with a trained SBERT model (Reimers, N. & Matthes, B, 2019). The idea behind semantic search is to find, for each job title, the semantically closest labels. “Closest” is defined by the cosine similarity between each vector of the labels and the vector of the job title to be classified. The label with the shortest distance, i.e., the closest label, is the predicted label for that class. As described in Chapter 2.1 the bkz label is semantically closer to the job titles than the KldB label. We therefore utilize the bkz labels to find the semantically closest label and infer, based on the bkz label id, the KldB label.

Figure 2 describes the process of the semantic search algorithm. First, all labels in the corpus are embedded using the trained BERT model. In this case, all labels of the bkz taxonomy. Then a query — in our case, the job title — is also embedded by the trained BERT model. Now both the labels and the query are in the same vector space. Then the cosine similarity between each label and the query can be calculated. The label with the highest cosine similarity is the closest label in that vector space and thus the most semantically similar label. This label is then to be chosen as the predicted label for the respective job title. The semantic search is performed using a pretrained BERT model (Vaswani, et al., 2017). The model and training procedure are described in detail in the following chapter.



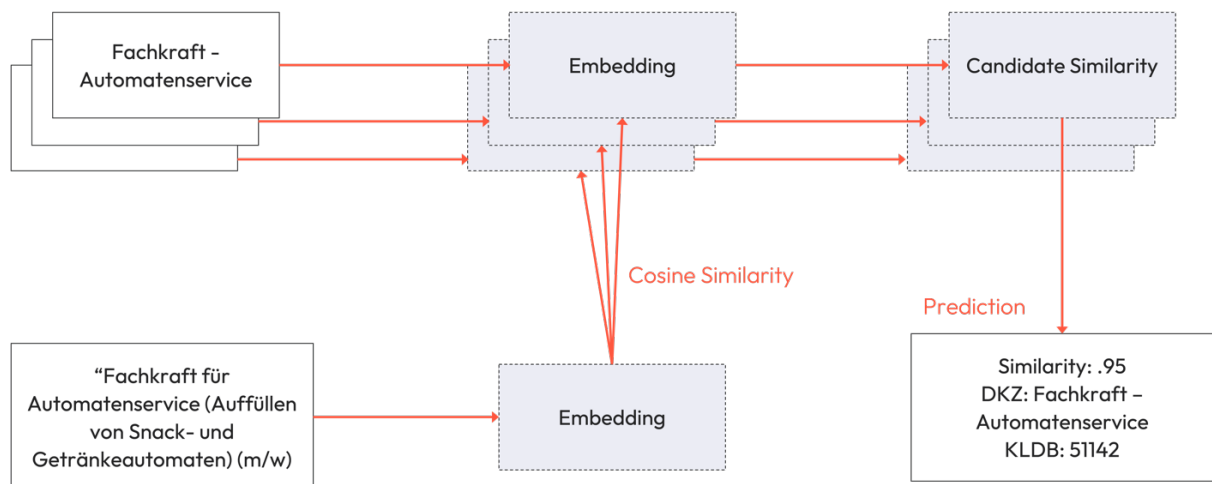


Figure 2: Semantic Search

## 3 Model

In the following, we describe the underlying extractor model of the semantic search following the model card framework approach of Mitchell, et al., 2018.

### 3.1 Model Details

The extraction model is a fine-tuned BERT model. The model is built on “bert-base-german-cased” (Deepset, 2020), which has 110 million parameters and is specifically trained for German. Each embedding has a vector length of 768 and a maximum sequence length of 256. The architecture of the model is composed of a word embedding layer and a pooling layer:

```

SentenceTransformer(
  (0): Transformer({'max_seq_length': 256, 'do_lower_case': False}) with Transformer model: BertModel
  (1): Pooling({'word_embedding_dimension': 768, 'pooling_mode_cls_token': False, 'pooling_mode_mean_tokens': True,
'pooling_mode_max_tokens': False, 'pooling_mode_mean_sqrt_len_tokens': False})
)

```

The trained model has a size of 436 MB. The speed at inference time on a 2,6 GHz 6-Core Intel Core i7 is 12.62 seconds for 1000 queries on the CPU and 1.35 seconds on the GPU, with an NVIDIA Tesla T4.

## 3.2 Intended Use

The trained BERT model can recognize similarities between occupations. As a Sentence Transformer model, the input of the raw model is a sentence. The output is an embedding. The embeddings can then be used to implement a classification model using semantic search and BKZ taxonomy as described in Chapter 2.7. The intended use of the model is to classify job titles from OJAs according to the KLDB taxonomy. The evaluation and optimization of the model is focused on this use.

## 3.3 Metrics

**Fine-Tuning Process.** For evaluating the fine-tuning process, we calculate Pearson's correlation and the Spearman rank correlation. The Pearson's correlation coefficient is calculated as follows:

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}}$$

with  $m_x$  and  $m_y$  defined as the mean of variables  $x$  and  $y$  (Virtanen 2020)

Pearson's correlation ranges from -1 to 1, with 1 indicating a positive monotonic association. It measures the linear relationship between the gold labels and the similarity between embeddings generated by the algorithm. Gold labels refer to the manually assigned similarity label for each pair we have created. To evaluate the results, the distance metrics cosine similarity, Euclidean distance, and Manhattan distance are used. However, since we use cosine similarity for the generation of the embeddings, the focus is on cosine similarity, and we expect a high positive correlation between the assigned similarity and the cosine similarity generated by the fine-tuned model for each pair.

The Spearman Rank Correlation is the ranked version of Pearson's correlation and is calculated using the rank of the metrics. It ranges as well between -1 and 1, and we expect a similarly high positive correlation (Chok, 2010).

**Classification.** For evaluating the semantic search and the classification of the job titles into the classes of the KLDB taxonomy, we use four metrics. We calculate the overall accuracy (OA) which measures how well the classification performs across all classes. Given  $tp_m$  are the correct prediction (true positives) for a current class  $m$ , the OA is the sum of the true positives over all classes divided by the number of all examples  $N$ :

$$OA = \frac{\sum_{i=1}^m tp_i}{N}$$

Further, we calculate macro precision, recall, and f1 score. The precision score indicates how well the classification detects the current class's actual correct predictions among all predictions for this class. The recall measures the identification of the current class's predictions among the actual correct predictions. F1 takes the harmonic mean between precision and recall. All three scores can be calculated as micro or macro scores. The macro score first calculates the respective metrics for each class separately against all other classes and then builds the average, while the micro score aggregates the values. Aggregation of the values leads to the same performance score for all four metrics. Thus, we only report the macro values.

Given that  $fp_i$  are examples that are falsely classified as the current class  $m$  and  $M$  is the total number of classes the macro precision is defined as follows:

$$precision_{macro} = \frac{1}{M} \sum_{i=1}^m \frac{tp_i}{tp_i + fp_i}$$

The recall is defined as:

$$recall_{macro} = \frac{1}{M} \sum_{i=1}^m \frac{tp_i}{tp_i + fn_i}$$

where  $fn_i$  are examples that are falsely classified not as the current class  $m$ .

The F1 Score is calculated by multiplying the  $precision_{macro}$  and the  $recall_{macro}$  score and dividing by the sum of both.

### 3.4 Training Data

We follow the Siamese Network Architecture (Reimers & Gurevych, 2019) approach to fine-tune the model. For this, we used the training dataset to build pairs of similar and dissimilar sentences and label them with a similarity score. For each sentence pair, each sentence is passed separately into the network, and an embedding is built. Based on the embeddings of each sentence, a similarity metric — in this case, the cosine similarity — is calculated. The calculated similarity score is compared to the labeled similarity score, and the weights of the model are updated accordingly.

We created similar and dissimilar pairs using the gold training dataset. Similar pairs are built with all job titles and their labeled BKZ names. Dissimilar pairs are defined as pairs of job titles with all BKZ names except the assigned BKZ name. Since the combination of dissimilar pairs is much higher, we used seven random BKZ names for each job title in the datasets. Similar pairs are labeled with 0.8, and dissimilar pairs with 0.2. In addition, for each BKZ name we excluded from the BKZ taxonomy as described in chapter 2.4, we calculated with a pre-trained model<sup>9</sup> the most similar BKZ name in the same KldB five-digit group based on cosine similarity. The most similar KldB name and the missing BKZ names were also passed as a pair of similar sentences, with a score of .95. We did this to make sure that, even if one of the excluded BKZ names in the taxonomy appears in a job title, the excluded BKZ name is still introduced over the most similar BKZ name for the same KldB five-digit group in the model. (sentence-transformers, 2019). For the training data set, we generated 21.308 similar pairs and 121.310 dissimilar pairs. In total, we used 142.618 pairs for the fine-tuning.

### 3.5 Evaluation Data

To evaluate the model during the fine-tuning and for the semantic search, we used 20% of the gold data for testing, referred to in the following as the "test dataset". To use the test data set to evaluate the fine-tuning, we processed the data the same way as the training data set. We generated 2.087 similar pairs and 11.746 dissimilar pairs.

### 3.6 Results Fine Tuning Process

The correlation metrics for evaluating the fine-tuning process are calculated for each training epoch. Table 4 shows, in general, a positive and high correlation for all correlation coefficients. However, Pearson's performance is much higher. This is probably because the similarity values are not normally distributed but are either close to 1 or close to 0. For non-normal data, the Spearman rank correlation is more resistant and better suited (De Winter, J.C. et al., 2016). The Pearson coefficient may overstate the correlation, which is why we report only the Spearman rank correlation for all distance metrics. Between the distance metrics, there is no notable difference. Moreover, the fine-tuned model's performance did not increase notably between the epochs.

Table 4: Evaluation of Fine Tuning

epoch	Euclidean Spearman	Manhattan Spearman	Cosine Spearman	Cosine Pearson
0	0.61	0.61	0.61	0.94

<sup>9</sup> We used the sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 (sentence-transformers, 2019). The model is evaluated for building qualitative sentence embeddings ([https://www.sbert.net/docs/pretrained\\_models.html](https://www.sbert.net/docs/pretrained_models.html))

1	0.62	0.61	0.62	0.96
2	0.62	0.62	0.62	0.96

Figure 3: Fine-Tuning Evaluation Results

### 3.7 Results Classification Task

Due to the systematic nature of the BKZ Taxonomy and the KLDB Taxonomy, it is possible to calculate the metrics on different levels. As stated, the most important levels are the first, third, and fifth levels of the KLDB ID. We calculate the metrics for each of these levels. In addition, since the direct predictions of the model are bkz names, we also report for each metric the BKZ id performance. In addition, since we are especially interested in the requirement level, we evaluated only the last digit against the true requirement level. Lastly, we concatenated the first, third, and fifth digits of each predicted and true KLDB id and calculated the metrics as well for those labels.

One problem with the macro scores is that if there is no prediction for a class, the macro score is zero for this class, and the precision is ill-defined. We, therefore, calculate, in addition, all macro scores, only including the classes that at least have one prediction. Thus, classes with zero predictions are not considered, and therefore those metrics have no explanatory power for all non-predicted labels. The true explanatory power of the model is therefore somewhere between the metrics of defined and ill-defined scores.

The following table shows the performance of the classification.

task	accuracy	precision (macro)	recall (macro)	f1 (macro)
KLDB LEVEL 5	0.86	0.72	0.71	0.70
KLDB LEVEL 3	0.89	0.81	0.83	0.80
KLDB LEVEL 1	0.95	0.94	0.93	0.93
BKZ ID	0.80	0.63	0.61	0.61
KLDB DIGIT 5 ONLY	0.94	0.92	0.93	0.93
KLDB LEVEL 5 (excluding ill-defined classes)	0.86	0.76	0.75	0.74
KLDB LEVEL 3 (excluding ill-defined classes)	0.89	0.82	0.84	0.81
KLDB LEVEL 1 (excluding ill-defined classes)	0.95	0.94	0.93	0.93

BKZ ID (excluding ill-defined classes)	0.80	0.71	0.69	0.68
KLDB DIGITS 1, 3, 5	0.88	0.77	0.77	0.76
COMBINED				

In general, the table shows a high performance for the classification. All performances exceed the requirements. Especially the high precision, recall, and f1 score of about 0.72% on Level 5 shows that even though the training data is imbalanced, the performance of the algorithms is still high, and the algorithm seems not to be biased. As expected, the performance gets lower with each level, since the number of classes is much higher at the bottom level. Of course, this is also because the performance at the upper levels determines the performance at the lower levels, as this is how the Kldb is structured. Excluding the zero predictions shows only significant differences on level five and on the BKZ ID level, indicating that there are more classes that are not predicted. The results in the range above 90% for the requirement level show that this can be differentiated very well by the algorithm and that cleaning the data by defining strict rules for the requirement level was beneficial. This becomes clear when comparing the evaluation of the requirement level of the baseline algorithms, which is about 80%. In general, comparing the performance of the algorithm to baseline performance shows an improvement, especially in the macro scores.<sup>10</sup> For example, for the Kldb ID on level five, the metrics improved by around 20%.

## 4 Application

We applied the algorithm to the data of Textkernel. The goal of the application is to enrich the Textkernel data with information on the Kldb occupation and the BKZ individual occupations based on the not-normalized job title using the algorithm instead of the mapping described in Chapter 2.6. We draw two samples from the Textkernel posting data. One random sample of 1000 examples from Textkernel is referred to as the "Textkernel random sample". For the second sample, we utilized the mapping based on O\*NET, which is already included in the Textkernel data, to get the top 50 demanded KldBs in January 2023, referred to as "Textkernel's top 50 sample".

---

<sup>10</sup> A table with all results of the baseline algorithm is provided in the Appendix. The evaluation was performed with the the test data as described in section 3.5 to ensure that the results are comparable.

First, we used the “Textkernel random sample” of the most demanded occupations using the BKZ names. The Figure 4 below shows the top 10 BKZ names for the Textkernel random sample along with the number of postings. You can see that the two individual occupations “Helfer/in – Lagerwirtschaft” and “Lager- und Transportarbeiter/in” in the warehouse management field are in the top ten most requested job titles. Accountants, as well as nursing professions, are also frequently in demand.

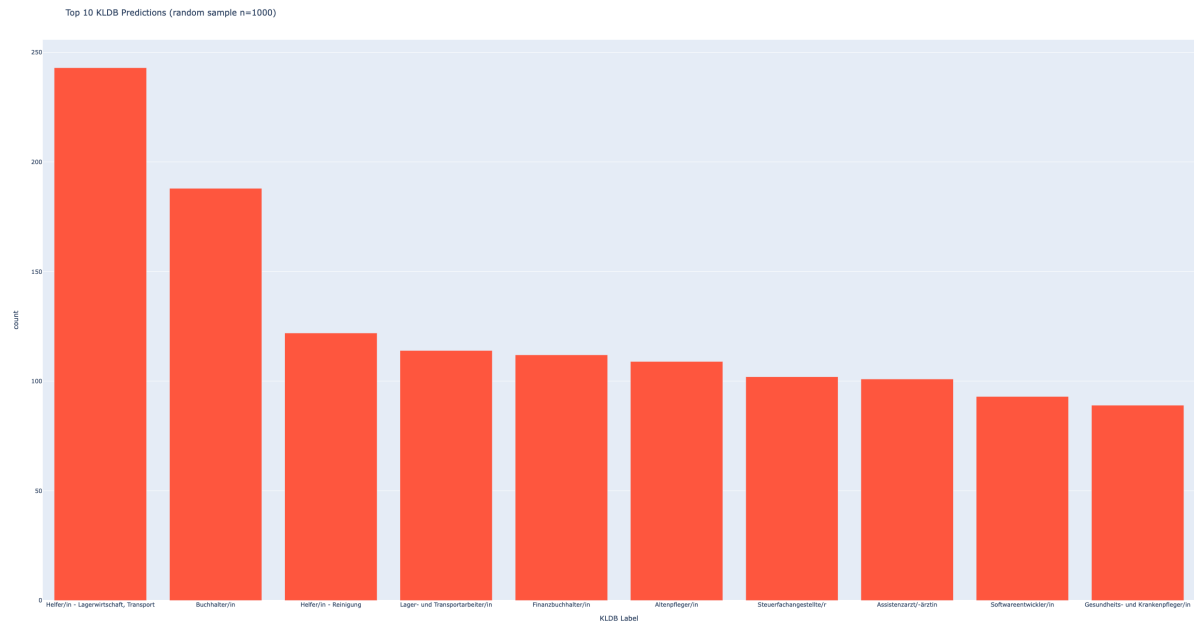


Figure 4: Top 10 KLDB Predictions

Second, we compared the KldB from the mapping to the KldB names we predicted with our algorithm to see whether there were major differences. We used “Textkernel's top 50 sample” for the analysis. Figure 5 shows the distribution of the Top 50 KldBs based on the mapping. To compare the mapping and the predictions of the developed algorithm, we predicted for all not-normalized job titles of the “Textkernel's top 50 sample” the KldB with the algorithm. The distribution of the predictions is shown in Figure 6. It can be seen from the two figures that the predicted KldBs for the job title increase. Thus, the algorithm allocates much more broadly. But it becomes also clear that, especially in the Top five, there are not many differences in the mapping, and the algorithm seems to assign similarly.

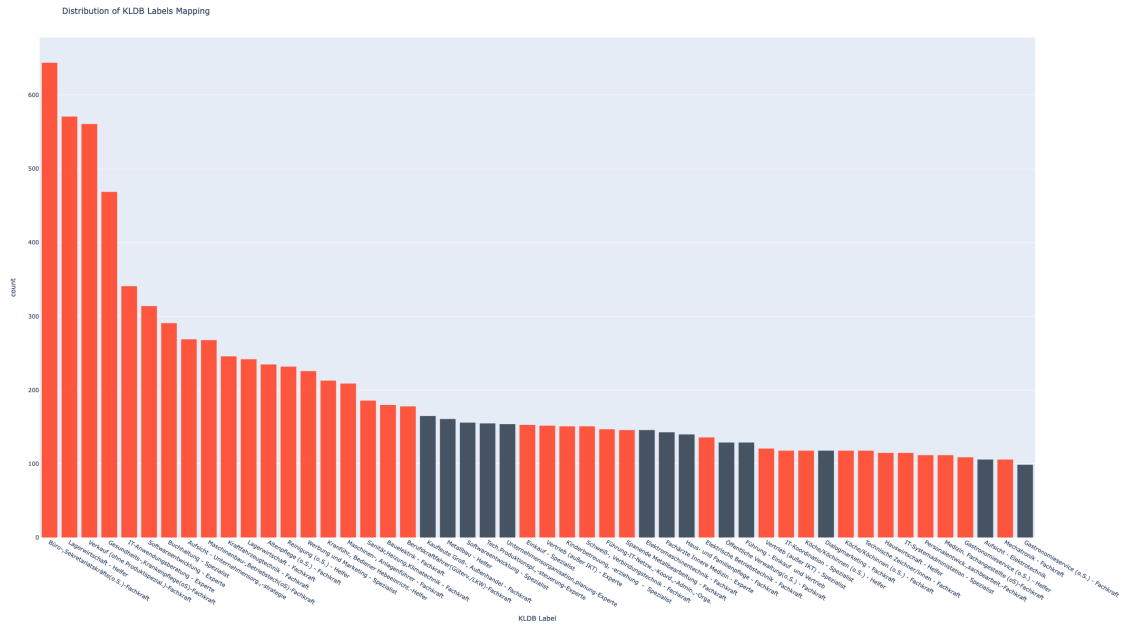


Figure 5: Distribution of KLDs of Top 50 Jobs assigned with mapping

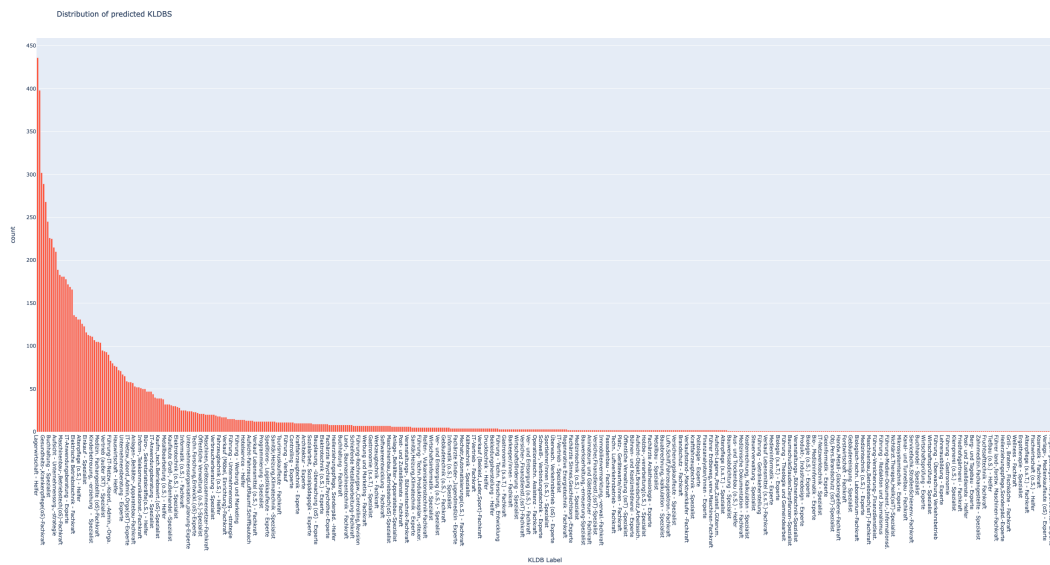


Figure 6: Distribution of KldBs of Top 50 Jobs assigned with algorithm

To further analyze the differences between the mapping and the predictions of the algorithm, we examine the KldBs of the mapping with a noticeable decline in the KldB names with respect to the predictions. Noticeable decline means here that the percentage decline in the number for the respective KldB between the old assignment based on the mapping and the new assignment with the algorithm is greater than 60%. The KldBs are colored in grey in Figure 5. For each grey colored KldB, we examined for the corresponding job titles which KldBs the algorithm predicts.



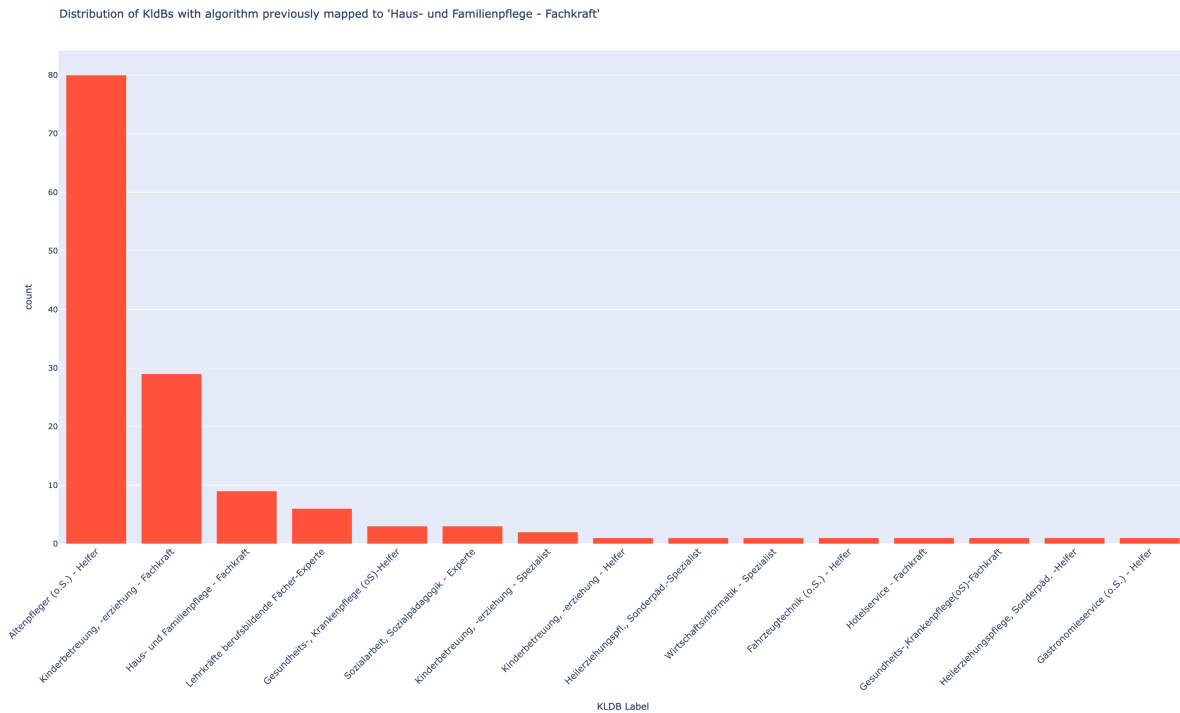


Figure 7: Distribution of titles mapped to "Haus-und Familienpflege-Fachkraft"

Figure 7 shows the prediction of the algorithm for all postings that are assigned to the KLDB "Haus- und Familienpflege – Fachkraft" according to the KLDB-Mapping. We can see that the assignment with the algorithm is much more differentiated. Many postings are now assigned to "Altenpflege (o.S.) – Helfer" and to "Kinderbetreuung, -erziehung – Fachkraft". Only nine out of the 140 posts assigned by the mapping are predicted by the algorithm to have the KLDB name "Haus- und Familienpflege – Fachkraft". In Table 5 we provided a snippet of the job titles assigned by the mapping to "Haus- und Familienpflege – Fachkraft". The table reveals, in fact, that the algorithm correctly predicts titles with the name "Altenpflege" as Altenpflege (o.S.) – Helfer" and titles like "Sozialpädagogische/r Assistent/in" or "Kinderpfleger" to Kinderbetreuung, -erziehung – Fachkraft. "Sozialpädagogische/r Assistent/in" is one of the bkz names that belong to the KLDB name "Kinderbetreuung, -erziehung – Fachkraft". Further, the last column shows that the mapping leads to wrong assignments, while the algorithm not only correctly assigns "Helfer Fahrzeugpflege ..." to the correct domain, but also recognize that the occupation is a requirement level 1 occupation. The analysis shows that assigning the KldB names with our algorithm differentiates much better than mapping the titles to the different KLDB names. Furthermore, since the algorithm is based on the bkz names, which are much closer in semantic terms to the job titles and do not rely on the mapping of O\*NET, it can predict those titles correctly, as seen in the example of "Sozialpädagogische/r Assistent/in".

Table 5: Comparison of Mapping and Algorithm Predictions

posting_id	job_title	old_kldb_mapping	new_kldb_prediction
1a4e94d2-6a46-44b2-acb3-eef9a33933d8	Altenpflegehelfer/in	Haus- und Familienpflege - Fachkraft	Altenpflege (o.S.) - Helfer
a472af73-8fda-410d-928f-3778cbe18ec0	Helfer/in - Altenpflege/Personliche Assistenz	Haus- und Familienpflege - Fachkraft	Altenpflege (o.S.) - Helfer
69d9e827-7846-4e58-8d53-9ef362920079	Pflegehelfer (m/w/d) - Gunzenhausen	Haus- und Familienpflege - Fachkraft	Altenpflege (o.S.) - Helfer
fe53dbab-3824-48ca-801b-c487dfa130b5	Sozialpädagogische Fachkraft	Haus- und Familienpflege - Fachkraft	Kinderbetreuung, -erziehung - Fachkraft
cfb22d69-5043-4e05-92c1-dd3f1ad55636	Helfer/in - stationäre Krankenpflege	Haus- und Familienpflege - Fachkraft	Gesundheits-, Krankenpflege (oS)-Helfer
219fc08c-c8d7-4891-9326-1eda90aa6578	Helfer Fahrzeugpflege (m/w/d) in Bentwisch	Haus- und Familienpflege - Fachkraft	Fahrzeugtechnik (o.S.) - Helfer

# Appendix

## A. Evaluation Rules

General Rules. This table contains all rules applying to all KldBs.

Table 6 (Appendix): General Evaluation Rules Data Cleaning

Rule	Example
If a bkz label is word for word in the job title, then all other bkz labels are wrong.	<b>Title:</b> "Bachelor of Science – Agrarwirtschaft" <b>assigned bkz label:</b> "Agrarwissenschaftler/in / Agrarökonom/in" <b>Evaluation:</b> 0 <b>Reason:</b> is not correct because the taxonomy contains a bkz label "Bachelor of Science – Agrarwirtschaft"
If a job title has the format "general occupation + detail", then all bkz labels with this general occupation need to be checked. If there is only one label and correctly assigned, then the title is evaluated with 1. If the bkz labels with the general occupation and specification do not have the same KldB id, then the specification must be checked. If there is no matching specification, then this title is evaluated with -999. It is also wrong if the bkz label of the format "general occupation" is used if there are bkz label with the format "general occupation + detail" for the respective occupation.	<b>Title:</b> "Zerspanungstechniker für CNC – Schleifentechnik" <b>assigned label:</b> "Zerspanungsmechaniker/in – Schleifentechnik" <b>Evaluation:</b> 1 <b>Reason:</b> correct because the specification of the title and the assigned label match.
If two bkz labels are word for word in the job title, then it is evaluated with 0.	<b>Title:</b> Schlosser/Industriemechaniker <b>Evaluation:</b> 0

If the job title does not contain an occupation it is evaluated with 0.

**Reason:** The bkz contains the bkz labels “Schlosser/in” and “Industriemechaniker/in”.

**Title:** “International führendes Unternehmen sucht Mitarbeiter”

**Evaluation:** 0

**Reason:** “Mitarbeiter” is not an occupation

If the job title has the format “general occupation” but there are only bkz label for this occupation in the format “general occupation + detail”, then this title is evaluated with 2.

**Title:** Verfahrensmechaniker/in

**Evaluation:** 2

**Reason:** There are many bkz labels like “Verfahrensmechaniker – Brillenoptik” or “Verfahrensmechaniker – Glastechnik” etc., but not “Verfahrensmechaniker”

If “Helfer”, “Fachhelfer” or “Produktionsmitarbeiter” is in the job title, then it must be a bkz id with requirement level 1.

**Title:** “Helfer Pferdewirtschaft”

**assigned bkz label:** “Pferdepfleger/in”

**Evaluation:** 0

**Reason:** Is not correct because the bkz id is 11302, which has the requirement level 2 and not 1.

Bkz Labels with “Ingenieur” are only allowed for job titles containing the word “Ingenieur” and vice versa. Evaluate with 0 if not.

**Title:** “Biotechnologe – Gruppenleiter, Qualitätskontrolle (m/w/d)”

**Assigned bkz label:** “Ingenieur/in – Biotechnologie”

**Evaluation:** 0

**Reason:** Ingenieur is not in title

If the word “Leiter” is in the job title the 4<sup>th</sup> digit of the labelled bkz id must be a nine. Also, if a job title is assigned to a bkz id with a 9 as the 4<sup>th</sup> digit, then the title is only evaluated with 1 if the word “Leitung” is explicitly in the title.

**Title:** Leiter – Automatisierungstechniker

**Assigned bkz label:** Ingenieur – Automatisierungstechniker

**Evaluation:** 0

If the title contains the word “Projektleiter” or “Projekmanager” then it must be assigned to one of the bkz labels with

**Title:** Projektleiter Verfahrenstechniker

**Assigned bkz label:** Projektleiter/in

**Evaluation:** 1

“Projektleiter/in” (with and without detail are allowed, except if the title specifically contains the “Weiterbildung/Ausbildung” or “Multimedia”

If the title contains the word “Projektingenieur” then it has to be labelled must be assigned to one of the bkz labels containing “Projektingenieur”. This has priority over bkz labels with “Ingenieur” in general

**Titel:** “Projektingenieur Verfahrenstechnik”

**Assigned bkz label:** Ingenieur/in – Verfahrenstechniker

**Evaluation:** 0

Projektmitarbeiter:in are labelled with - 999

KldB specific rules. Additionally, to the general evaluation rules, we have formulated for each KldB specific rules. This also might lead to conflict the general rules, because depending on the KldB some general rules might not fit. Specific rules have priority over general rules.

Table 7 (Appendix): Specific Evaluation Rules Data Cleaning

## KLDB Rule

1	“Landwirtschaftlicher Mitarbeiter” is correct if it is assigned to the requirement level “Helfer- und Anlern Tätigkeiten”
1	The bkz label “Helfer/in – Gartenbau” is only correct if the title solely is about Gartenbau. Otherwise, the bkz label “Helfer/in – Land-, Forst-, Tierwirtschaft, Gartenbau” is correct. Same applies for Landschaftsgärtner with the bkz labels “Gärtner – Garten und Landschaftsbau” und “Landschaftsgärtner/in”
2	Job titles with “Fenstermonteur” are only evaluated with 1 if they are assigned to the bkz label “Glaser/in – Glasbau/Fensterbau”
2	Job titles with “Laminierer” are evaluated with 2, except if the job title is specified with “Kunststoffverarbeitung”, then evaluate only with 1 if the bkz label is “Laminierer/in (Kunststoffverarbeitung)”

2	Job titles with “Küchenmonteur” are only evaluated with 1 if they are assigned to the bkz label “Fackraft Möbel-Küchen und Umzugsservice”.
2	“Schreiner” is the same as “Tischler”.
2	Job titles with “Hilfskraft Verpackung” are only evaluated with 1 if they are assigned to the bkz label “Helfer/in Papier, Verpackung”
2	“CNC Fachkraft” is not the same as “CNC Programmierer”. If in same bkz label, then label with 0.
2	“Pulverbeschichtung” is equal to “Oberflächen”
2	Job titles of the occupation chemical engineer without any specification are evaluated with 2.
2	Mechatroniker and Mechaniker are not the same
2	Job titles with “Gerätemonteur” are only evaluated with 1 if they are assigned to the bkz label “Gerätezusammensetzer/in”
2	Job title with “Montagemitarbeiter” are evaluated with 1 if they are assigned to the bkz label “Montierer/in”
2	Machschinenbediener and Maschineneführer are the same
2	Titles containing the both “Maschinenführer” and “Anlagenführer” are only evaluated with 1 if they are assigned to the bkz label “Maschinen- und Anlagenführer/in (o. Angabe des Schwerpunkts)”
2	Job titles with “Serviceelektriker” are evaluated as -999.
2	<p>Steps for evaluating Servicetechniker if a bkz label with “techniker” instead of “servicetechniker” is assigned</p> <ol style="list-style-type: none"> <li>1. First check all bkz labels with “Servicetechniker”. If there is a matching label, then evaluate with 0.</li> <li>2. If no matching label, then check all bkz labels with “Techniker” and “Monteur”. If the assigned labels matches, then evaluate with 1.</li> <li>3. If again no matching, evaluate with -999</li> </ol>
2	Job titles with “Instandhalter:in” are evaluated with -999
2	Job titles with “Fahrzeugfolierer” are only evaluated with 1 if the assigned bkz label is “Fahrzeuglackierer/in”
2	Job titles with “Kfz-Monteur” are only evaluated with 1 if they are assigned to the bkz label “Fahrzeugschlosser/in”
2	Job titles with “KFZ-Aufbereiter (m/w/d) are evaluated with -999

- 2 Job titles with “Schwergeratemechaniker:in” are only evaluated with 1 if they assigned to the bkz label “Kraftfahrzeugmechatroniker/in - Nutzfahrzeugtechnik”
- 2 Job titles with “KFZ-Kundendiensttechniker:in” are only evaluated with 1 if they assigned to the bkz label “Kraftfahrzeugmechatroniker/in - Personenkraftwagentchnik”
- 2 Job titles with “KFZ-Kundendiensttechniker:in” are only evaluated with 1 if they assigned to the bkz label “Kraftfahrzeugmechatroniker/in - Personenkraftwagentchnik”
- 2 Job titles containing certifications for helicopters, airplanes or similar are evaluated with -999
- 2 “Betriebsmechatroniker” and “Mechatroniker” are equal
- 2 Job titles like “Mechatroniker (m/w/d) als Servicemonteuer” are evaluated with 0
- 2 Job titles with “Elektrokonstrukteure” are evaluated with 1 only if they are assigned to the bkz label “Elektroniker/in - Automatisierungstechnik”
- 2 Job titles with “Elektronikfachkraft” are evaluated with -999
- 2 Job titles with “SPS-Programmierer” is evaluated with 1 only if they are assigned to the bkz label “SPS-Fachkraft”
- 2 Job titles with “Automatisierungstechniker” are evaluated with 1 if the bkz-label is “Fachkraft - Automatisierungstechnik”
- 2 Job titles with “Betriebselektriker” are evaluated with 1 if they are assigned to the bkz label Elektroniker/in – Betriebstechnik”
- 2 If the job title “Elektriker” is assigned to bkz label “Elektroinstallateur”, then it is evaluated as 0.
- 2 Job titles with “Elektromitarbeiter” are evaluated as -999
- 2 Job titles with “Lagerverwalter/in” are evaluated as -999
- 2 “Elektriker” and “Elektroniker” are not the same.
- 2 Job titles like “Elektrofachkraft für Gebäudetechnik” or other detail are evaluated with 1 if they are assigned to the bkz label “Elektrofaachkraft für festgelegte Tätigkeiten”  
Job titles with “Fachkraft für den Maschinenbau” are evaluated with -999
- 2 The bkz label “Elektroniker/in - Masch. u. Antriebstechnik (Ausbildung bis 2021)” is evaluated as wrong, expect “Ausbildung bis 2021” is mentioned in the title.

- 2 Job titles like “MSR Techniker” are evaluated with 1 if assigned to bkz label “Mess-techniker/in (Energietechnik)”
- 2 Job titles with “Technischer Produktmanager Windenergie” are evaluated with 1 if assigned to bkz label “Produktmanager/in”
- 2 “Anlagemechatroniker:in” is not the same as “Elektroanlagenmonteur:in”
- 2 Job titles with “Schaltbauschrank” are evaluated with 1 if assigned to bkz label “Elektroniker/in - Energie- und Gebäudetechnik”
- 2 Job titles with “Büroinformationselektroniker/in” are evaluated with -999
- 2 “Bachelor of Engineering” is equal to “Ingenieur/in”
- 2 “Elektrotechnikmeister:in” is equal to “Elektromeister:in”
- 2 “Informationstechniker” is equal to “Informationstechnikermeister/in”
- 2 Job titles with “CAD” are evaluated with 1 if assigned to the bkz label “CAD-Fachkraft (ohne Schwerpunkt)”
- 2 Job titles with “Arbeitsvorbereiter” are evaluated with 1 if assigned to the bkz label “Arbeitsvorbereiter/in” except if the job titles contains the word “Rechenzentrum”, then only evaluate with 1 if bkz label is “Arbeitsvobereiter/in (Rechenzentrum)”
- 2 Job titles with “Qualitätsplaner” are evaluated with -999
- 2 Job titles with “Fertigungsprüfer” are evaluated with -999
- 2 Ingenieur has priority over lead position for Qualitätsingenieur.
- 2 There is no clear difference between “Produktionsleiter/in” and “Leiter/in - Produktion/Fertigung“. Matching titles assigned to “Produktionsleiter/in” are evaluated with 1, else 0.
- 2 “Qualitätsmanagement” has priority over “Teamleitung”
- 2 Job titles with “Textilverarbeitung Anlernetätigkeit” are evaluated with -999
- 2 Job titles with “Schaumstoffkonfektionist” are evaluated with 1 if assigned to bkz label “Kunststoff- u. Schwerebekonfektionär/in”
- 2 The bkz label “Näher/in (§66 BBiG/\$42m HWO) is only evaluated with 1 if the job title explicitly contains the §66
- 2 Job titles with “Produktionshelfer:in” for “Wurstherstellung” or “Schokoladenproduktion” are only evaluated with 1 if assigned to “Helfer Lebensmittelherstellung”
- 2 Job titles with “Essensausgabenkraft (m/w) are only evaluated with 1 if assigned to “Servicekraft – Gastronomie”



- 2 Job titles about “Helfer Küche” are only evaluated with 1 if assigned to “Helfer/in Küche”
- 2 Job titles with “Spülhilfe” are only evaluated with 1 if assigned to “Helfer/in Küche”
- 2 Job titles with “Küchenkraft” and “Küchenmitarbeiter” are equal to job titles with “Helfer Küche.”
- 2 Job titles with “Alleinkoch/Alleinköchin”, “Spezialitätenkoch/Spezialitätenköchin”, “Sushikoch”, “Teilköchin” and similar title are only evaluated with 1 if assigned to bkz label “Koch/Köchin”
- 2 “Küchenleiter:in” and “Küchenmeister:in” are equal to “Küchenchef”
- 2 If job titles with “Objektleiter für Restaurant” are assigned to bkz label “Küchenchef” they evaluated as 0.
- 3 Bkz labels with “Bautechniker” are equal to bkz labels with “Techniker/in – Bautechnik”. Matching titles assigned to one of those bkz labels are evaluated with 1
- 3 Bkz labels with “Ingenieur/in – Bau” are equal to bkz labels with “Ingennieur/in Bau (allgemeines Bauwesen)”. Matching titles assigned to one of those bkz labels are evaluated with 1.
- 3 Job titles with “Experte Baubetrieb” are evaluated with 0 if assigned bkz label is “Fachkraft”
- 3 Job titles with “Dipl.Ing.” for “Straßenplanung/Verkehrsbau” are evaluated with 1 if assigned bkz label is “Verkehrsbauingenieur/in”
- 3 “Bauoberleiter” and “Bauüberwacher” is equal to “Bauleiter/in”
- 3 “Bauleiter” has priority over “Projektleiter/in”
- 3 The bkz label “Bauhilfsarbeiter/in” is evaluated for all job titles with 0.
- 3 The bkz label “Hochbaufachwerker/in” is evaluated for all job titles with 0.
- 3 Job titles with “Eisenflechter:in” are evaluated with 1 if assigned to bkz label “Bet-onstahlbieger/in und -flechter/in”
- 3 “Helfer Rohrleitungsbau” are equal to “Helfer/in im Tiefbau”
- 3 The bkz label “Tiefbautechniker/in (Kultur und Wasserbau)” is only evaluated with 1 if “Kultur und Wasserbau” are explicitly in the job title
- 3 Job titles with “Bauingenieur für Tiefbau- und Medienplanung” are evaluated with 0

- 3 Job titles with “Fliesenleger/in” are evaluated with 1 if assigned to the bkz label “Fliesen-, Platten- und Mosaikleger/in”
- 3 Job titles with “Helfer Maler:in” are only evaluated with 1 if assigned to the bkz label “Helfer/in – Malerei, Lackierei”
- 3 If job title is “Maler:in” it is evaluated with 1 if assigned to the bkz label “Maler/in und Lackierer/in” – Maler/in
- 3 Bkz label “Maler/in u. Lackierer/in” is evaluated always with 0.
- 3 Job titles with “Gipser” are evaluated with 1 if assigned to bkz label “Stuckateur/in
- 3 Job titles with “Trockenbauer” are evaluated with 1 if assigned to bkz label “Fachkraft – Trockenbau”
- 3 “Verkaufsberater” has priority over “Holzbau/Baustoff”
- 3 The bkz label “Bautischler/in” is only evaluated with 1 if “Bautischler” is explicitly in the job title
- 3 “Gebäudetechniker/in” is the same as “Hausmeister/in”
- 3 Job titles with “Helfer HLS” or “Helfer Sanitär” are evaluated with 1 if the assigned bkz label is “Helfer/in – Klempnerei, Installation”
- 3 Job titles with “Heizungsbauer:in” or “Lüftungsbauer:in” are evaluated with 1 if the assigned bkz label is “Zentralheizungs- und Lüftungsbauer/in”
- 3 Job titles with “Heizungsmonteur:in”, “Heizungstechniker:in”, “Installateur für Sanitärtechnik” or “Kundendienst für Sanitär” are evaluated with 1 if the assigned bkz label is “Anlagenmechaniker/in – Sanität-, Heizungs- und Klimatechnik”
- 3 Job titles with “Müllwerker:in” are evaluated with 0.
- 3 Job titles with “Wasserwerker:in” are evaluated with 1 if the assigned bkz label is “Fachkraft – Wasserversorgungstechnik”
- 3 Job titles with “Kanalreiniger” or “Rohhreininiger” are evaluated with 1 if the assigned bkz label is “Fachkraft – Rohr-, Kanal- und Industrieservice”
- 3 A “Vorrichter” is equal to “Rohrvorrichter”
- 3 Job titles with “Behälterbauer” are evaluated with 1 if the assigned bkz label is “Behälter- und Apparatenbauer”
- 4 Job titles with “Mathematiker” are evaluated with 1 if the assigned bkz label is “Mathematiker/in” even if there is specification. Exception: Specification matches literally one of the bkz label specification like “Wirtschaftsmathematiker”

- 4 In “scientific” occupations English titles are allowed and if correct assignment then is labeled with 1.
- 4 Job title “Naturwissenschaftler:in” is evaluated as zero.
- 4 “Fachlehrer/in – allgemeinbildende Schulen” has priority over “Biologielehrer/in”
- 4 Bkz label “Biologielaborant” only if explicitly something with “Biology” is in the title.
- 4 “Probennehmer” only correct if labelled with bkz label “Probennehmer/in (Met-  
allerzeugung)
- 4 Job titles with “Pharmawerker/in” are labelled with 1 if assigned to bkz label “Phar-  
makant/in”
- 4 Job titles with “Kundenbetreuer/in in Energiebranche” are evaluated with 0 if bkz  
name “Gebäudenergieberater/in” is assigned.
- 4 Job titles with “Ausbildung Fachinformatiker” is evaluated with 0
- 4 Job titles with “Service Desk Support” or “Fachinformatiker” are evaluated with 1 if  
assigned to the bkz label “Fachinformatiker/in – Systemintegration”
- 4 Job titles with “ERP” are evaluated with 1 only if a bkz label with “ERP” is assigned
- 4 Job titles with “Softwaretester” are evaluated with 1 only if bkz label is “IT-Tester.
- 4 BKZ label “Assistent/in – Informatik” is only evaluated with 1 if job title contains the  
word “Assistent”
- 4 Bkz label with “Entwickler/in Digitale Medien” and “IT-Systemplaner/in” are only  
evaluated with 1 if explicitly in job title.
- 4 Job titles with “Fullstack Developer” are only evaluated with 1 if assigned to BKZ la-  
bel “Softwareentwickler/in”
- 5 Job titles with “Brandposten” are evaluated with 1 if assigned to bkz label “Sicher-  
ungsposten”
- 5 Job titles with “Lagermitarbeiter” are evaluated with 1 if assigned to bkz label “La-  
ger- und Transportmitarbeiter/in”
- 5 Job titles with “Helfer und Lager”, “Helfer und Kommissionierer” or “Versand und  
Helfer” are evaluated with 1 if assigned to bkz label “Helfer/in – Lager, Versand,  
Transport, Verkehr”
- 5 Job titles with “Warenverräumer” or “Waren” are evaluated with 0 if assigned to  
bkz label “Kommissionierer”
- 5 Job titles with “Zustellung Zeitung” are evaluated with 1 if assigned to “Helfer/in –  
Kurier, Zusteller und Postdienstleitungen”

5	Job titles with “Postmitarbeiter:in” or “Zusteller:in” or “Postbote:in” are evaluated with 1 if assigned to bkz label “Fachkraft – Kurier-, Express- und Postdienstleistungen”
5	Job titles like “Bürokraft für Taxibetriebe” are evaluated with 0 if assigned to bkz label “Funbediener/in (Taxibetrieb)”
5	Job titles containing “Logistikplaner, Materialplaner are evaluated with 1 if assigned to bkz label “Logistiker/in”
5	Job titles containing “Kraftfahrer:in” or have a specification like “Kraftfahrer/in C/CE 12-40t zur Auslieferung von Baustoffen” are evaluated with 1 if assigned to the bkz label “Berufskraftfahrer/in”
5	Job titles containing “LKW Fahrer” are evaluated with 1 if assigned to bkz label “Berufskraftfahrer/in”
5	Job titles containing “Kurierfahrer:in” or “Paketzusteller” are evaluated with 1 if assigned to bkz label “Auslieferungsfahrer/in (nicht Verkaufsfahrer/in)”
5	Bkz labels “Baumaschinenführer/in”, “Baugeräteführer/in“ or “Kraftfahrzeugführer/in von Baufahrzeugen” are only evaluated with 1 if job title contains the labels literally
5	Job titles containing “Radladefahrer:in ” are evaluated with 1 if assigned to bkz label “Baumaschinenführer/in”
5	Job titles like “Monteur für Brandschutz” are evaluated with 1 if assigned to bkz label “Servicetechniker/in – Sicherheits Alarmanlagen”
5	Job titles containing “Reinigungskraft” are evaluated with 1 if assigned to bkz label “Helfer/in – Reinigung”
5	Job titles containing “Raumpfleger” are evaluated with 1 if assigned to bkz label “Helfer/in – Reinigung”
5	Job titles containing “Glas und Gebäudereiniger:in” are evaluated with 1 if assigned to bkz label “Glasreiniger/in” or to bkz label “Gebäudereiniger/in”
5	Job titles containing “Unterhaltsreiniger:in” are evaluated with 1 if assigned to bkz label “Gebäudereiniger/in”
5	Job titles containing “Fahrzeugaufbereiter:in” are evaluated with 0 if assigned to bkz label “Fahrzeugpfleger/in”
6	Job titles containing “Einkäufer:in” are evaluated with 1 if assigned to a bkz label with “Einkäufer/in” + specification. If job titles contain “technischer Einkäufer:in”,

	then they are only evaluated with 1 if assigned to bkz label “Technischer/in Einkäufer/in”
6	Job titles assigned to the bkz label “Außendienstmitarbeiter/in” are only evaluated with 1 if assigned bkz label is literally in the job title.
6	Job titles assigned to the bkz label “Fachkraft – Marketing/Verkauf/Vertrieb” are evaluated with 0
6	Job titles containing “Vertriebsprofi” or “Vertriebsexperte:in” are evaluated with 1 if assigned to bkz label “Vertriebsberater/in”
6	Job titles containing “Innendienst” or “Vertriebsinnendienst” are evaluated with 1 if assigned to bkz label “Verkaufssacharbeiter/in”
6	Job titles containing “Vertrieb Leiter:in” are evaluated with 1 if assigned to bkz label “Leiter/in – Vertrieb” regardless of the specification in the job titles
6	Job titles containing “Regionalleiter:in” are evaluated with 1 if assigned to bkz label “Bezirksleiter:in”.
6	Job titles containing “eCommerce and Shop-Manager/Manager” are evaluated with 1 if assigned to bkz label “Kaufmann/-frau E-Commerce”
6	Job titles containing “Immobilienverwalter:in” are evaluated with 1 if assigned to bkz label “Objektverwalter/in (Immobilien)”
6	Job titles containing “Property Manager:in” or “Betriebskostenabrechner:in” are evaluated with 1 if assigned to bkz label “Immobilienkaufmann/-frau”
6	Job titles assigned to the bkz label “Fachwirt/in – Immobilien” are evaluated with 1 if Job titles contains “Immobilienfachkraft:in” or “Fachwirt:In Immobilien”
6	Job titles assigned to the bkz label “Betriebswirt/in (Fachschule) – Immobilien” are evaluated with 0
6	Job titles assigned to the bkz label “Fachwirt/in – Facility-Management” are evaluated with 0.
6	Job titles containing “Warenveräumer:in” are evaluated with 1 if assigned bkz label is “Helfer/in – Verkauf”
6	Job titles containing “Regalauffüller:in” are evaluated with 0 if assigned to bkz label “Helfer/in – Verkauf”
6	Job titles containing “Empfangskraft” are evaluated with 1 if assigned to bkz label “Empfangskraft”

6	Job titles containing "Reservierungsmitarbeiter:in" are evaluated with 1 if assigned bkz label is "Reservierungssachbearbeiter/in (Hotel)"
6	Job titles containing "Helfer:in Spülkraft" are evaluated with 0 if assigned to bkz label "Helfer/in - Gastgewerbe"
6	Job titles containing "Servicekraft Gastgewerbe" are evaluated with 0 if assigned to bkz label "Servicekraft – Gastronomie und Gastgewerbe"
7	Job titles containing "Büromanagement and Kaufmann/-frau" are evaluated with 1 if assigned to bkz label "Kaufmann/-frau Büromanagement". The same rule applies for "Industriekaufmann/-frau"
7	Job titles containing "Projektkoordinator:in" are evaluated with 1 if assigned to bkz label "Projektleiter/in"
7	Job titles containing "Teamleiter:in" or "Gruppenleiter:in" are evaluated with 1 if assigned to bkz label "Gruppen-, Teamleiter/in"
7	Job titles with "Standortleiter:in" or "Gebietsleiter:in" are evaluated with 1 if assigned to bkz label "Niederlassungsleiter/in"
7	Job titles containing "Notfall" and are assigned to bkz label "Telefonist/-in" are evaluated with 0.
7	Job titles containing "Bürokraft" or "Büromitarbeiter:in" are evaluated with 1 if assigned to bkz label "Bürokaufmann/-frau"
7	Job titles containing "Office Manager:in" are evaluated with 1 if assigned to bkz label "Kaufmann/frau – Büromanagement"
7	Job titles assigned to bkz label "Bürogehilf(e/in)" are only evaluated with 1 if job titles contain word for word "Bürogehilf(e:in)"
7	Job titles with "Schreibkraft" are evaluated with 0.
7	Job titles containing "Teamassistent:in" are assigned to bkz label "Teamassistentz"
7	Job titles containing "Debitoren-/Kreditorenbuchhalter:in" are evaluated with 1 if assigned to bkz label "Finanzbuchhalter/in"
7	Job titles containing "Lohnspezialist:in", "Lohnexperte:in" or similar are evaluated with 1 if assigned to bkz label "Lohn- und Gehaltsbuchhalter/in"
7	Job title containing "Assistenz" are only evaluated with 1 if assigned bkz label contains "Assistenz" as well
7	Job title containing "Volljurist" are evaluated with 1 if assigned bkz label is "Jurist/in"

7	Job titles containing “Rechtsanw(alt:ältin)” + specification are evaluated with 1 if assigned bkz label is “Rechtsanwalt/Rechtsanwältin”
8	“Pflegeassistent”, “Pflegehelfer” and “Altenpfleger” are equal.
8	Job titles containing “Pflegehilfskräfte” are evaluated with 1 if assigned to “Altenpflegerhelfer/in”
8	Job titles with “Arzthelfer/in” and Job titles with “MFA” are not equal and have to be assigned to correctly to BKZ label “Arzthelfer/in” or “Medizinische/r Fachangestellte/r”
8	Job titles with “Prophylaxeassistent/in” are evaluated with label 1 if assigned to “Zahnmedizinische/r Prophylaxeassistent/in”
8	Job titles with “Kinderkrankenpfleger” are evaluated with label 1 if assigned to “Gesundheits- und Kinderkrankenpfleger/in”
8	Job titles “Intensivpfleger/in” are evaluated with label 1 if assigned to “Fachkrankenpfleger/in - Intensivpflege/Anästhesie”
8	Job titles with “Gesundheits- und Krankenpfleger” are evaluated with label 0 if not assigned to bkz label “Gesundheits- und Krankenpfleger/in”
8	Job titles with “Wohnbereichsleitung” are evaluated with 0 if assigned to bkz label “Stationsleiter/in – Kranken-/Alten-/Kinderkrankenpflege”
8	Job titles with “Arzt in Weiterbildung” are evaluated with 1 if assigned to bkz label “Arzt/Ärztin”
8	Job titles with “Salonleitung” are evaluated with 1 if assigned to bkz label “Friseurmeister/in”
8	Job title with “Erzieher:in” only evaluated with 1 if assigned bkz label is “Erzieher/in”
8	Job titles with “Hauswirtschaftler” only evaluated with 1 if assigned bkz label is “Hauswirtschaftler/in (ohne Schwerpunkt)”
8	Job titles with “Lehrer/-in” without any specification are evaluated with 1 if bkz label Lehrer/in (Uni) - Gesamtschulen”
8	Nachhilfkräfte are evaluated only with 1 if assigned to bkz label “Lernbetreuer/in”

## B. Recoding Rules

In this section for each specific word in the title, like “Sachbearbeiter” recoding rules are listed.

## Sachbearbeiter:in

Table 8(Appendix): Rules Sachbearbeiter:in

Job titles with	BKZ
Sacharbeiter Buchhaltung	Buchhalter/in
Kaufmännischer Sachbearbeiter in der Buchhaltung	Fachkaufmann/-frau - Geschäfts- u. Finanzbuchführung
Sachbearbeiter Geschäfts- und Anlagenbuchhaltung	Finanzbuchhalter/in
Sachbearbeiter (m/w/d) - Lohnbuchhaltung	Lohn- und Gehaltsbuchhalter/in
Sacharbeiter Abrechnung	
Sacharbeiter Entgeltabrechnung	
Sachbearbeiter (m/w/d) Finanzbuchhaltung	Finanzbuchhalter/in
Sachbearbeiter Debitoren..	
Personalsacharbeiter	Personalsacharbeiter/in
Sacharbeiter HR	
Sacharbeiter Bewerbung	
Sacharbeiter Kreditsacharbeiter	Kreditsacharbeiter/in
Sacharbeiter Logistik	Sachbearbeiter/in - Logistik
Sacharbeiter Versand	
Sachbearbeiter Produktionsplanung	
Sachbearbeiter Supply Chain Sachbearbeiter Disposition Sachbearbeiter Fertigungssteuerung Sachbearbeiter Materialwirtschaft	
Sacharbeiter Luftfracht	Sachbearbeiter/in - Luftaufsicht
Sacharbeiter Einkauf	Sacharbeiter/in - Einkauf
Sachbearbeiter Kundenbetreuung und Kundenservice (wenn nicht fachspezifisch)	



Sacharbeiter Auftragsbearbeitung/abwicklung

Sachbearbeiter Beschaffung Sachbearbeiter Forderungsmangement

Sacharbeiter Verkauf

Sachbearbeiter Vertrieb(-innendienst)

Sachbearbeiter Ersatzteil Sachbearbeiter aufträge alle Spezifizierungen

Sachbeabeiter Reklamation

Sachbearbeiter Kreditor/Kredit

kaufmännische Sachbearbeiter ohne Spezifizierung

Bürosachbearbeiter

Sacharbeiter Export/Import/Zoll/Spedition/Transport

Sacharbeiter Immobilien, Miete, Haus und

Kaufmännischer Sachbearbeiter im Personal

Sacharbeiter Rechnungswesen

Sacharbeiter Heizungs-, Lüftungs-, Klimatechnik

Landschaftspflege

Sacharbeiter Insolvenz

IT Sacharbeiter

kaufm. Sachbearbeitung Transport

kaufm. Sachbearbeitung Vertrieb

kaufm. Sachbearbeitung Elektro

kaufm. Sachbearbeitung Rechnungswesen

kaufm. Sachbearbeiter technisch

kaufm. Sachbearbeiter Logistik

Verkaufssachbearbeiter/in

Reklamationssachbearbeiter/in

Kreditsachbearbeiter/in

Sacharbeiter/in - Einkauf → nicht klassifizierbar

Bürokaufmann/-frau

Speditionskaufmann/-frau

Immobilienkaufmann/-frau

Fachkaufmann/-frau - Personal

Assistent/in Fachkraft - Rechnungswesen

Techniker/in - Heizungs-, Lüftungs-, Klimatechnik

Natur- und Landschaftspfleger/in

Rechtsanwaltsfachangestellte/r

IT-Systemadministrator/in

Speditionskaufmann/-frau

Fachkaufmann/-frau Vertrieb

Kaufmann/-frau - Einzelhandel (Elektrogeräte)

Kaufmann/-frau - Geschäfts- u. Finanzbuchführung

Technische/r Kaufmann/-frau

Kaufmann/-frau - Spedition und Logistikdienstleistung

Sachbearbeitung - Medizinische Schreibkraft	Medical Writer
Sacharbeiter in Kombination mit BTA, CTA, PTA	Rangfolge: BTA, CTA, PTA (entsprechend was dabei steht)
Marketing	Fachkaufmann/-frau - Marketing
Sachbearbeiter/in Sale	Sales-Manager/in
Sachbearbeiter Vertrag, Rechtsabteilung	Justiziar
...	
Sacharbeiter Versicherungen wie Altersversorgung	Versicherungssachbearbeiter/in
Sacharbeiter Controlling	Assistenz Controlling
Sachbearbeiter (m/w/d) Energie(wirtschaft)	Fachwirt/in Energiewirtschaft
Sacharbeiter Notrufservices/Notfall ...	Sicherheitsmitarbeiter/in
Sachbearbeiter Backoffice	Bürokaufmann/-frau
Sozialpsychiatrischen Dienst, Sozialer Dienst Betreuung	Sozialarbeiter/in / Sozialpädagoge/-pädagogin
Sachbearbeiter Kinder- und Jugendärztlicher Dienst	Facharzt/in - Kinder- und Jugendmedizin
Sachbearbeiter Post	Kaufmann/-frau - Kurier, -Express- u. Postdienstleistungen
Sachbearbeiter Revision	Revisor/in
Technischer Sacharbeiter Kommunikation	Assistent/in - Technische Kommunikation
Sacharbeiter Vermessung	Techniker/in - Vermessungstechnik
Sacharbeiter + BKZ z.B. Sachbearbeiter Arbeitsvorbereitung	dann diese BKZ zuordnen Arbeitsvorbereiter/in
Sachbearbeiter Bank	Dokumentensachbearbeiter/in (Bank)
Sachbearbeiter + Kontrolle	Baukontrolleur/in (Sicherheitskontrolleur/in)
Sachbearbeiter Objektverwaltung	Objektverwalter/in (Immobilien)
Sachbearbeiter Gebäudemangement	
Sacharbeiter Gewässer	Fachkraft - Wasserwirtschaft
Sachbearbeiter Kundendialog	Kundendienstberater/in

Sachbearbeiter Leitstelle	Funkstellenleiter/in (Rettungsleitstelle)
Sachbearbeiter technische Dokumenta- tion	Technische/r Redakteur/in
Sachbearbeiter Automation	Fachkraft - Automatisierungstechnik
Sachbearbeiter Stadterneuerun/planung etc.	Stadt- und Regionalplaner/in
Schadenssachbearbeiter	Schadensregulierer/in (Versicherung)
Sachbearbeiter TGA	Techniker/in - Versorgungstechnik

### Ingenieur:in

Table 9 (Appendix): Rules Ingenieur:in

Job titles with	Bkz
General Rule: For a job title, if there is a matching “Ingenieur” with specification in the bkz taxonomy, then it will be recoded to this bkz label. For example: The job title “Prozessingenieur Optik/Laser” will be recoded to “Ingenieur/in – Lasertechnik”	
Bau	
if Hochbau	Ingenieur - Bau Hochbauingenieur
if Tiefbau	Ingenieur/in - Tiefbau
if Wasserbau	Wasserbauingenieur/in
if Wege	Wegebauingenieur/in
if Straßenbau	Straßenbauingenieur/in
With the rule “Bauingenieur:in” has priority over “Architekt:in” and “Bauleiter:in” and no further specification	
Entwicklungsingenieur	Ingenieur - Bau
Same rule applies for “Projektingenieur/in” and “Berechnungsingenieur/in”	Forschungs- und Entwicklungsingenieur/in
Garten -und Landschaftsbau	Ingenieur/in - Landschaftsarchitektur
Verpackung, Pack...	Ingenieur/in - Verpackungstechnik
Holz	Ingenieur/in - Holztechnik

Qualität, Validierung	Qualitätsingenieur/in
Versuch, Test, Prüf	Testingenieur/in
Anlagen und Anlagenbau	Ingenieur - Automatisierungstechnik
Steuerungstechnik	Ingenieur - Automatisierungstechnik
Kraftwerkstechnik	Ingenieur/in - Elektrotechnik
Batteriemodulfertigung	Ingenieur/in - Elektrotechnik
Leittechnik	Ingenieur/in - Elektrotechnik
Vakuumtechnik	Ingenieur/in - Physik
Kerntechnik	Ingenieur/in - Physik
Fertigungsprozess	Ingenieur/in - Verfahrenstechnik
Logistik	Vertriebsingenieur/in
MSR	Ingenieur/in - Elektrotechnik
Management	Wirtschaftsingenieur/in

### Referent:in

Table 10 (Appendix): Rules Referent:in

#### **Job titles with**

#### **Bkz**

Generally tried to recode to requirement level 3 or 4 positioner (5th position), manager (but not if fourth position 9), consultant etc. preferred.

Gesundheit, Klinik, Arzt, medizinisch      Referent/in - klinische Studien (CRA)

Controlling

Controller/in

Bilanzierung

Bilanzbuchhalter/in

Finanzen, Anlagen

Finanzbuchhalter/in

Recht, Jurist

Jurist/in

Compliance

Compliance-Manager/in

Account

Account-Manager/in

Rehabilitation

Rehabilitationspädagoge/-pädagogin

Vetrieb, Einkauf/Beschaffung

Fachkraft - Marketing/Verkauf/Vertrieb

Organisationsentwicklung,

Unternehmensberater/-in

Prozessentwicklung

Versicherung

Versicherungsberater/-in

### Assistent:in

Table 11 (Appendix): Rules Assistent:in

Job titles with	BKZ
Teamassistent	Büroassistent
Recht	Rechtsanwaltsfachangestellter/r
Technische Assistenten if Naturwissen- schaften	BTA

### Consultant:in

Table 12 (Appendix): Rules Consultant:in

Job titles with	BKZ
IT-Consultant Alles IT und Data related	IT-Berater/in
Consultant company Strategie Geschäft	Unternehmensberater/in
Account	Fachberater/in - Finanzdienstleistungen
Salesforce/sale	CRM-Consultant
Marketing etc.	Unternehmenberater/in
Pflanzen	Kundeberater/in - Gartenbau
Immobilien	Berater/in - Immobilienfinanzierung
Logistik/Vetrieb...	Vertriebsberater/in
Küchenfachberater	Küchenplaner/in (Ausnahme Berater!)
Kundenberatung Telefon/Telefonberatung	Telefonist/in (Ausnahme Berater!)
Tele-Sales Tele-Marketing	
Energieberater	Berater/in - Erneuerbare Energien
Verkauf	Verkaufsberater/-in
Migration	Interkulturelle/r Kommunikationsberater/in

### Further Rules

- Job titles with “Pflegekraft” are recoded to “Pflegefachmann/-frau”.
- Job titles with bkz “Eisenbahner/in – Betrieb. – Lok. u. Transp. (Ausb. Bis 2022)” are recoded to Eisenbahner/in – Betriebsdienst – Lokführer/in und Transport” or “Rangierbegleiter/in”

- Job titles with “BTA”, “CTA”, “MTA”, “PTA”: Since almost all job titles contain all three 3 BKZ Labels or other BZK Labels such as “Chemiker/in” or “Biologe/in”. We therefore define the following priority order: “BTA”, “CTA”, “PTA”, “MTA” for recoding. If job title is combined with other bkz labels like “Biologe” this must be considered.
- Job titles combining “Biologie” with other science subjects are all recoded to “Biologe/Biologin”
- Job titles with “Vertriebsdisponent:in” are recoded to “Fachkaufmann/-frau - Vertrieb”
- Job titles containing “Elektrotechniker:in” are recoded to “Techniker/-in – Elektrotechnik” (depends on specification), otherwise -999
- Job titles with “Gebietsleiter:in” are recoded to “Gebietsleiter/in – Außenwirtschaft”
- Job titles with “Elektriker:in – Betriebstechniker” are recoded to “Industrieelektriker/in – Betriebstechnik”
- Job titles with “IT-Spezialist” are recoded to “EDV-Fachkraft”
- Job titles with “kaufmännische Mitarbeiter:in” are recoded to “Kaufmännische Fachkraft”
- Job titles with “Lehrkraft – Biologe” are recoded to “Lehrer/in – Gymnasien”
- Job titles with “Vertriebsprofi” or “Vertriebsexperte:in” are recoded to “Vertriebsberater/in”
- Job titles with “Informatiker:in” are recoded to “Informatiker/-in (Hochschule)”. It is not relevant which other science subject is given in the title.
- Job titles with “SAP-Anwendung” are recoded to “ERP-Systembetreuer-in”.
- Job titles with “IT-Mitarbeiter/in” are recoded to “Fachinformatiker/in – Systemintegration”.
- Job titles with “IT-Techniker” are recoded to “Techniker/in – Informatik (technische Informatik)
- Job titles with “IT Digitalisierung” are recoded to “Digital-Analyst/in”
- Job titles with “KYC ” are recoded to “Business-Analyst/in”
- Job titles with “Manager” and “IT” are recoded to “IT-Manager/in
- Job titles with “Vertriebler” and “Außendienst” are recoded to “Außendienstmitarbeiter/in”
- Job titles with “Verwaltung” + “Amt” are recoded to “Verwaltungsfachangestellte/r – Kommunalverwaltung”
- Job titles with “IT-Cloud” are recoded to “Cloud-Architect”.

- Job titles with “IT – Application Manager” are recoded to “Application-Engineer/-Manager/in”.
- Job title with “IT Rollout” are recoded to “Servicetechniker/in – IT”.
- Job titles with “IT help desk, support” are recoded to “IT-Berater/in”.
- Job titles with “Financial analyst” or “Financial Controller” are recoded to “Controller/-in”
- Job titles with “Custom Care” or “Kundenservice” are recoded to “Kundendienstberater/-in”
- Job titles with “Kundenbetreuer:in” (also by phone) are recoded to “Kundeninformationsfachkraft”
- Job titles with “Fachanwalt:in” + specification are recoded to “Fachanwalt/-anwältin”
- Job titles with “payroll” are recoded to “Lohn und Gehaltsbuchhalter/in”
- Job titles with “Software Engineer” are recoded to “Softwareentwickler/in”
- Job titles with “Sozialdienstmitarbeiter/-in” are recoded to “Sozialarbeiter/in / Sozialpädagoge/-pädagogin“
- Job titles with „Vertriebsmitarbeiter/in“ are recoded to „Vertriebsassistent/in“

## Baseline Evaluation Results

The evaluation was performed with the completely cleaned dataset to ensure that the results are comparable.

Table 13 (Appendix): Baseline Evaluation Results

<b>task</b>	<b>accuracy</b>	<b>precision (macro)</b>	<b>recall (macro)</b>	<b>f1 (macro)</b>
kldb_id_level_5	0.67	0.53	0.51	0.49
kldb_id_level_3	0.81	0.67	0.66	0.65
kldb_id_level_1	0.90	0.84	0.84	0.84
bkz_id	0.57	0.41	0.37	0.38
kldb_level_5_last_digit	0.82	0.80	0.79	0.79

# Literature

- Baskaran, R. (2022). Job Title Classification - A Comparison of Vectorization and Classification Techniques for German Job Postings. Universität Konstanz.
- Bundesagentur für Arbeit (2021). Klassifikation der Berufe 2010 - überarbeitete Fassung 2020 Band 1: Systematischer und alphabetischer Teil mit Erläuterungen.
- Bundesagentur für Arbeit (2023a). Download Portal Arbeitsagentur <<https://download-portal.arbeitsagentur.de>> (last seen 01.09.2023).
- Bundesagentur für Arbeit (2023b). Jobsuche <https://www.arbeitsagentur.de/jobsuche/> (last seen 01.09.2023).
- Chok, N.S. (2010). Pearson's Versus Spearman's and Kendall's Correlation Coefficients for Continuous Data. University of Pittsburgh.
- Deepset. German Bert. (2020). <https://huggingface.co/bert-base-german-cased> (last seen 01.09.2023).
- De Winter, J. C., Gosling, S.D. & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data, *Psychological methods*, 21(3), 273.
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the conference on fairness, accountability, and transparency* (pp. 220-229).
- O\*NET OnLine. National Center for O\*NET Development <<https://www.onetonline.org/>> (last seen 07.11.2023)
- Paulus, W., & Matthes, B. (2013). Klassifikation der Berufe - Struktur, Codierung und Umsteigeschlüssel. Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Reimers, N. & Matthes, B. (2019). Sentence-Bert - Sentence embeddings using siamese bert networks. arXiv.
- SBERT. 2023. Semantic Search. <https://www.sbert.net/examples/applications/semantic-search/README.html> (last seen 01.09.2023)
- Textkernel. (2022, 16. September). Textkernel. < <https://www.textkernel.com/de/>> (last seen 01.09.2023).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.



Vicari, B. (2014). Grad der standardisierten Zertifizierung des Berufs – Ein Indikator zu Messung institutioneller Eigenschaften von Berufen (KldB 2010, KldB 1988). Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, 17(3), 261-272.

## Über &effect

Unser Ziel ist es, Data Science zu einem integralen Bestandteil der Entscheidungsfindung im öffentlichen und sozialen Sektor zu machen. Dafür entwickeln wir wirkungsorientierte Datenprodukte an der Schnittstelle zwischen Sozialwissenschaften, Data Science und Softwareentwicklung.

### Kontakt:

[info@and-effect.com](mailto:info@and-effect.com)

### Herausgeber:

&effect data solutions GmbH

Pasteurstraße 34

10407 Berlin