

Jobmonitor.de: Datenaufbereitung, Anreicherung und Berechnung der Metriken

Methodenbericht

Projekt: Jobmonitor

Bertelsmann Stiftung

Version

2.0.0

Autoren

Johannes Müller, Gunvald Herdin & Hauke Janson
(Bertelsmann Stiftung)

Erstellt am

03. September 2022

Letzte Änderung am

25. Juni 2026

&effect

Inhalt

1	Zusammenfassung	3
2	Anreicherung	4
2.1	Berufsbezeichnungen	4
2.2	Transversale Kompetenzen.....	7
2.3	Teilqualifikationen	8
2.4	Landkreise und kreisfreie Städte nach der NUTS-Klassifikation.....	9
3	Datenaufbereitung	10
3.1	Deduplizierung	10
3.2	Bereinigung von Stellenausschreibungen in der Zeitarbeit	11
3.3	Aggregation	14
4	Metriken	14
4.1	Berufe	14
4.2	Soft Skills (transversale Kompetenzen)	15
4.3	Teilqualifikationen	16
5	Verweise.....	16
6	Tabellenverzeichnis.....	17

1 Zusammenfassung

Online-Stellenausschreibungen bieten ein großes Potenzial für die Analyse von Trends im Arbeitsmarkt. Um Ergebnisse einem breiten Publikum zugänglich zu machen, wird der Jobmonitor.de entwickelt – ein Dashboard, das monatsaktuelle Daten zu Online-Stellenausschreibungen aufarbeitet. Der Jobmonitor macht Veränderungen am Arbeitsmarkt erstmalig regional und tagesaktuell sichtbar. Durch Auswertung von über 54 Millionen Online-Stellenanzeigen aus ca. 300 Millionen Postings können Detailanalysen für etwa 1.000 Berufe, 95 Soft Skills und 400 Kreise und kreisfreie Städte erstellt werden. Regionale Arbeitsmarktakteure können ihre Maßnahmen mit den Analysen des Jobmonitors gezielt auf die aktuellen Bedarfe ausrichten.

Das vorliegende Papier beschreibt, wie die Daten angereichert, dedupliziert und aggregiert wurden. Außerdem wird beschrieben, welche Metriken auf dem Jobmonitor.de dargestellt werden. Grundlage für die Analysen des Jobmonitors ist eine Datenbank von mittlerweile über 400 Mio. Online Stellenanzeigen aus über 50.000 Quellen (z.B. Jobportale, Firmen-Webseiten, Zeitungen). Die Daten stammen vom niederländischen Anbieter Textkernel BV, der seit dem Jahr 2014 Daten in Deutschland erhebt und aufbereitet. Alle 350 Millionen Stellenausschreibungen werden in einer Datenbank der Bertelsmann Stiftung (DOSTA) abgelegt.

Update Version 2 (Juli 2026):

- Neues Modell (v3) für die Extraktion von Berufen (KLDB-2010) das auf alle Jahre angewendet wird (2018-)
- Neue Berechnung des Zeitarbeitgewichtes
- Filterung von Praktikumsstellen

Für den Jobmonitor werden die Daten in mehreren Schritten angereichert und aggregiert: 1) Mapping der Berufe aus der TextKernel-Taxonomie auf die Klassifikation der Berufe 2010 (KldB-2010); 2) Extraktion von transversalen Kompetenzen nach ESCO v1.1 aus den Volltexten; 3) Anreicherung um NUTS-3 Region über geografische Daten der Ausschreibung; 4) Deduplizierung und Aggregation der Stellenanzeigen; 5) Berechnung der relevanten Metriken.

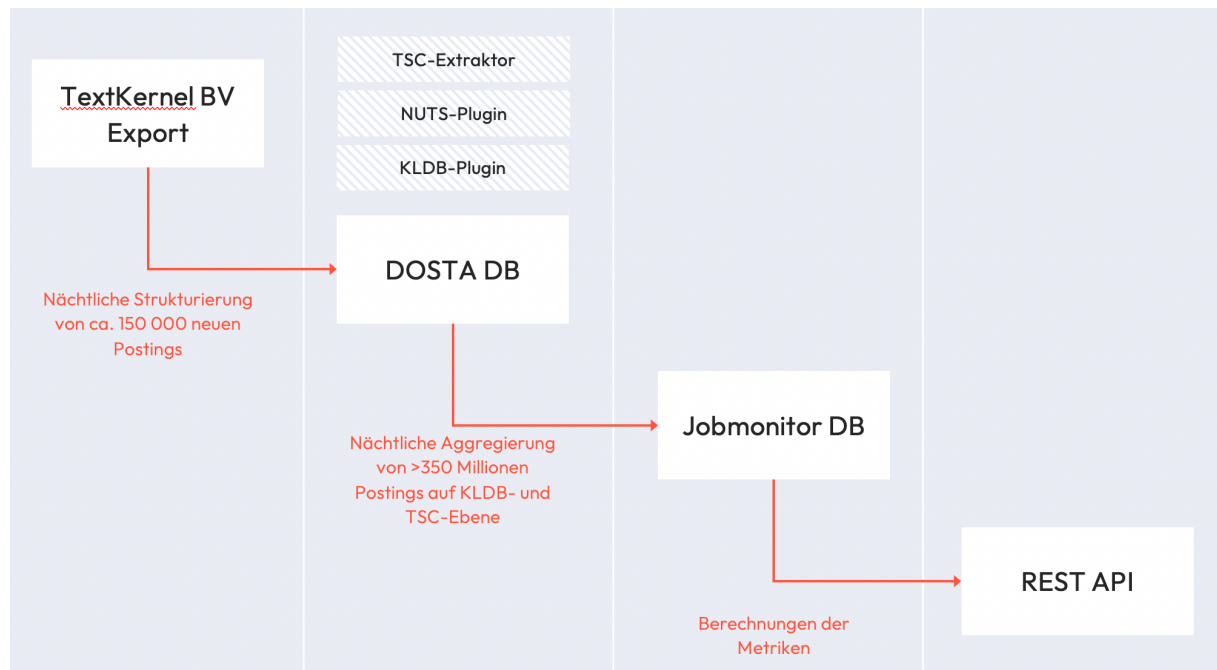


Abbildung 1 Aggregation und Anreicherung der Online-Stellenanzeigen.

2 Anreicherung

Online-Jobausschreibungen weisen neben dem Freitext der Ausschreibung diverse strukturierte Meta-Daten auf. Für den Jobmonitor.de sind insbesondere vier Variablen relevant: das Veröffentlichungsdatum, der Ort des ausschreibenden Unternehmens, zugeordnete transversale Kompetenzen und die Berufsbezeichnung nach der Klassifikation der Berufe 2010 (KldB-2010).

2.1 Berufsbezeichnungen

Um Online-Stellenausschreibungen einen Beruf zuordnen zu können, müssen die Berufsbezeichnungen systematisiert und gruppiert werden. Der Jobmonitor benutzt dafür die Klassifikation der Berufe 2010 (KldB-2010). Die KldB-2010 ist die offizielle Taxonomie der Berufe in Deutschland – sie wird unter anderem von der Bundesagentur für Arbeit und dem statistischen Bundesamt genutzt.

Taxonomie: Die Klassifikation der Berufe wurde entwickelt, um die Berufslandschaft in Deutschland zu systematisieren. Die KLDB-2010 ist eine hierarchische Klassifikation mit fünf Gliederungsebenen. Dabei werden Berufe sowohl nach einer horizontalen Dimension (Berufsfachlichkeit) und einer vertikalen Dimension (Anforderungsniveau) strukturiert.

Die Berufe werden mithilfe eines fünf-stelligen Codes numerisch kodiert (Tabelle 2). Die ersten vier Ebenen gruppieren Berufe anhand ihrer Berufsfachlichkeit und die unterste Ebene beinhaltet das Anforderungsniveau.

Tabelle 1 Überblick Klassifikation der Berufe 2010

Stelle	Ebene	Anzahl der Kategorien
1	Berufsbereich	10
2	Berufshauptgruppe	37
3	Berufsgruppe	144
4	Berufsuntergruppe	700
5	Berufsgattung	1286

Dem KLDB-5-Steller sind konkrete Berufsbezeichnungen zugeordnet. Diese werden in der Datenbank der DKZ (Dokumentationskennziffern), dem zentralen Datenbankregister der Bundesagentur für Arbeit, laufend aktualisiert. Die Systematik der Einzelberufe folgt dabei der KldB-2010 und kann über eine eindeutige ID-Nummer, der DKZ-8-Steller, den entsprechenden KLDB-5-Steller zugeordnet werden (Vicari 2014).

Anreicherung: Für die Extraktion von Berufsbezeichnungen aus einer Online-Stellenausschreibung sind zwei Schritte notwendig. In einem ersten Schritt muss die freie Berufsbezeichnung (z.B. die Überschrift einer Stelle) korrekt extrahiert werden. Im zweiten Schritt muss diese Berufsbezeichnung mit einer Taxonomie verglichen und der richtige Beruf zugeordnet werden.

Für die Analyse von Berufen nach KldB-2010 müssen Berufsbezeichnungen dementsprechend zugeordnet sein. Die Entwicklung eines Extraktionsalgorithmus hat dabei eigene Herausforderungen, die umfangreich adressiert werden müssen. Insbesondere die Differenzierung nach Berufsuntergruppen und Berufsgattung unterscheidet die KldB-2010-Taxonomie von anderen Taxonomien, wie zum Beispiel ISCO-08 oder der TextKernel-Taxonomie. Weiterhin ist die Taxonomie mit insgesamt 1286 Kategorien im Vergleich zu ISCO-08 (ca. 400 Kategorien) relativ groß.

Für die Extraktion und wurde im Juli 2023 ein neuer Algorithmus implementiert (v2). Eine ausführliche Beschreibung der Methoden und der Evaluationsergebnisse sind im Methodenbericht (Baskaran und Müller 2023) zu finden.

Neues Berufsextraktionsmodell (v3) ab Juli 2026

Im Jahr 2026 fand ein weiteres Update des Modells statt mit der Version 3: Unter anderem wurden das Extraktionsmodell um aktuelle Berufsbezeichnungen ergänzt und auf einem breiteren Trainingsdatensatz trainiert um eine bessere Abdeckung zu gewährleisten (Müller 2026). Relevant für den Jobmonitor ist außerdem, dass ein Beruf hinzugefügt wurde, der außerhalb der KLDB-2010 Taxonomie liegt. Der KLDB-5-Steller „Produktionshelfer/in (allgemein)“ wurde eingeführt um eine große Anzahl an Stellen für „Produktionsmitarbeiter/innen“ und „Produktionshelfer/innen“ einzuordnen, die nicht klar einem bestimmten Helferberuf zuzuordnen sind. Der neue Beruf erhält die Platzhalter-ID „20001“ und gliedert sich in die Taxonomie unter der Berufsgruppe „Produktionsberufe (allgemein)“ (200) und dem bestehenden Berufsbereich „Rohstoffgewinnung, Produktion, Fertigung“ (2) ein.

Daraus ergeben sich ein Abweichungen im Vergleich zu Metriken die auf Basis der v2 berechnet wurden. Hier der Vergleich der Top-10-Berufe zwischen v2 (aus 2023) und v3 (aus 2026). Bei den absoluten Zahlen gilt es zu beachten, dass hier bereits das neue Zeitarbeitsgewicht und der Filter der Praktikumsstellen verwendet wird.

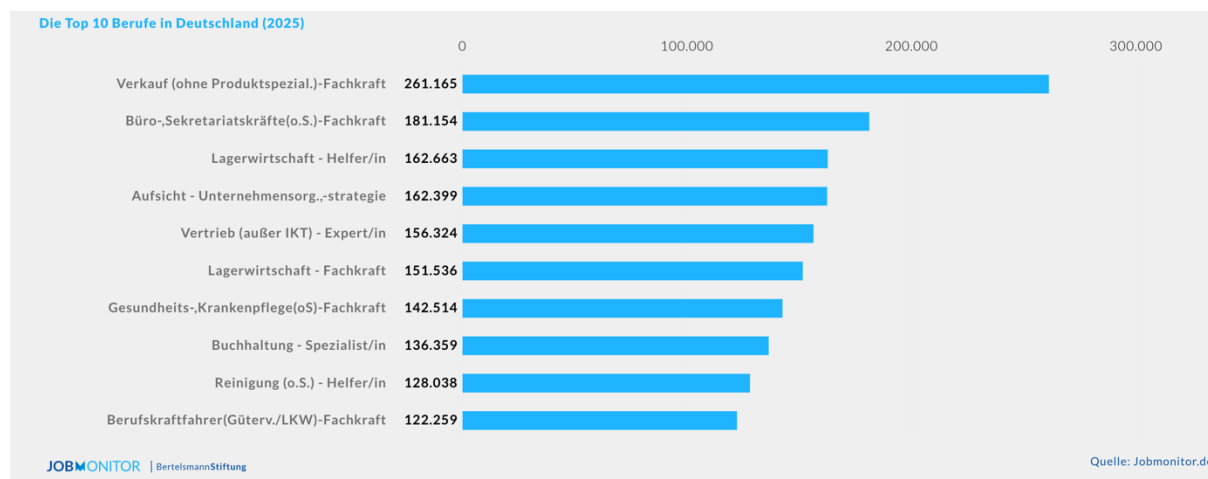


Abbildung 2 Top 10 Berufe mit Filterversion v1 und Berufsextraktionsmodell v2

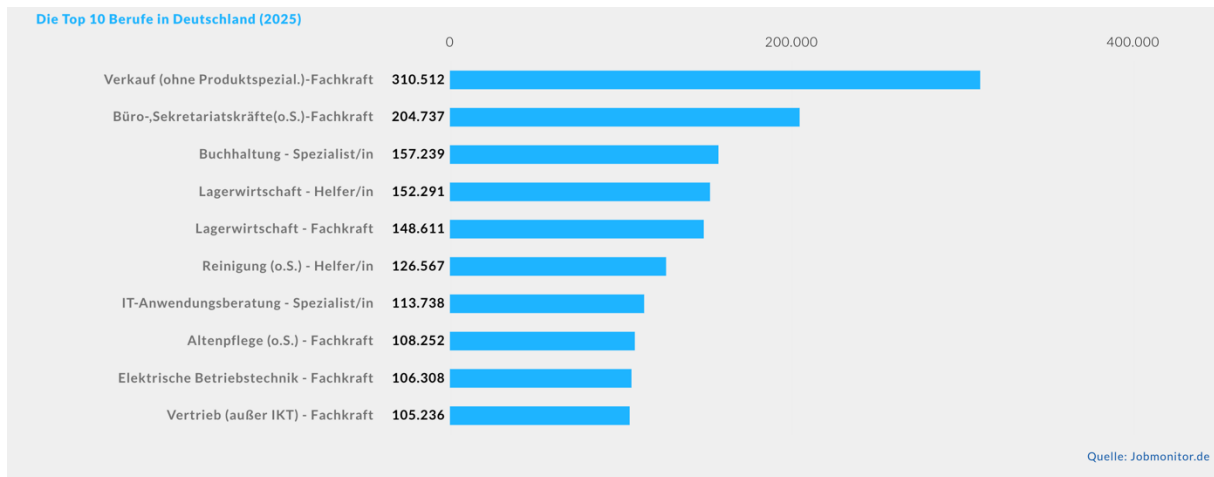


Abbildung 3 Top 10 Berufe mit Filterversion v2 und Berufsextraktionsmodell v3

2.2 Transversale Kompetenzen

Hinweis: Die Extraktion von transversalen Kompetenzen nach ESCO v1.1 wird an dieser Stelle nur zusammengefasst für eine ausführliche Besprechung des Modells verweisen wir auf den Methodenbericht (Müller et al. 2022).

Für die strukturierte Analyse von Online-Jobanzeigen sind Kompetenzen wertvolle Konzepte. Neben den berufsbezogenen Kompetenzen sind insbesondere die transversalen Kompetenzen relevant. Transversale Kompetenzen sind definiert als Kompetenzen, die im Allgemeinen als wichtig für Arbeits-, Lern- oder Lebensaktivitäten erachtet werden und nicht spezifisch für einen Beruf oder eine bestimmte Tätigkeit sind. Hierunter fallen z.B. Einsatzbereitschaft, Teamfähigkeit, Sprachkenntnisse und Umweltbewusstsein.

Taxonomie: Es existieren verschiedene Taxonomien transversaler Kompetenzen. Eine besonders vollständige, mehrsprachige und frei zugängliche bietet ESCO v1.1 von der Europäischen Kommission (Ziegler, 2022). Die Bertelsmann Stiftung und die &effect data solutions GmbH entwickelten und evaluierten das erste Modell für die Extraktion von transversalen Kompetenzen nach ESCO v1.1 in deutschen Online-Jobanzeigen. Für das Extraktionsmodell wird eine adaptierte Version von ESCO v1.1 genutzt¹.

¹ Die ESCO-Kompetenz „Informationen und ihre Quellen kritisch bewerten“ wird als Teil von „kritisch denken“ behandelt und nicht als eigenständige Kompetenz. Statt 69 werden in der Adaption 189 Sprachkompetenzen erfasst, wobei „Alt-Griechisch“ und „Belarussisch“ in ESCO v1.1 vorkommen, jedoch nicht in der vorliegenden Adaption. In 16 Kategorien und Clustern wurde ein zusätzliches Konzept „sonstige“ mit eigenen Suchwörtern erstellt (z.B. „sonstige soziale und kommunikative Kompetenzen“ oder „sonstige Digitale Grundkompetenzen“). Dies geschah, um relevanten Suchwörtern, die keinem der bestehenden Konzepte zuzuordnen waren, aber eindeutig zur nächsthöheren Ebene gehörten eine Verortung im Modell zu geben. Konzeptbezeichnungen wurden in vielen Fällen durch am deutschen Arbeitsmarkt gängigere Begriffe ersetzt.

Anreicherung und Evaluation: Die Basis des Extraktionsmodells stellt eine Ontologie dar. Diese enthält insgesamt 16176 Suchwörter für 95 transversale Kompetenzen und weitere 189 Sprachkompetenzen. In insgesamt sechs Annotationsrunden wurden die Suchwörter um Synonyme, Lexikalisierungen und kontextualisierte Formen der Kompetenzen erweitert. Für jede Runde annotierten Expert:innen Stellenanzeigen mit Kompetenzen, um die Modellgüte zu evaluieren. Die Annotationsübereinstimmung für den finalen Goldkorpus ist mit einem Fleiss' Kappa von .80 auf Kompetenz-Ebene, .86 auf Cluster-Ebene und .89 auf Kategorie-Ebene sehr gut.

Das finale Modell besteht aus einem regelbasierten Named-Entity-Recognition-Modell, das Wortsequenzen aus Online-Jobanzeigen extrahiert. Über alle Kompetenzen hinweg (Micro-Scores) hat das Modell auf Dokument-Ebene eine Precision von .88, einen Recall von .83 und einen F1-Score von .85. Die leicht niedrigeren Macro-Scores mit einer Precision von .77, Recall von .70 weisen auf eine unterschiedliche Performanz zwischen den Konzepten hin. Insbesondere für seltene Kompetenzen (z. B. Sprachkompetenzen abgesehen von Deutsch und Englisch) kann das Modell deshalb in Zukunft noch weiterentwickelt werden. Wenn das Modell nicht auf segmentierten Stellenanzeigen, sondern auf Volltexten angewendet wird, hat das Modell eine leicht schlechtere Precision, aber einen leicht verbesserten Recall.

2.3 Teilqualifikationen

Hinweis: Die Extraktion von Teilqualifikationen wird an dieser Stelle nur zusammengefasst für eine ausführliche Besprechung des Modells verweisen wir auf den Methodenbericht (Müller 2023).

Extraktion: Als Extraktionsmodell für Teilqualifikationen wurde für jede Berufsgruppe ein eigenes Modell trainiert (insgesamt 27). Jedes Modell besteht aus zwei Komponenten: einem regelbasierten und einem statistischen Wortsequenz-Erkennungs-Algorithmus. Beide Komponenten schlagen Wortsequenzen in Stellenausschreibungen vor, die sich auf ein oder mehrere Tätigkeitsfelder beziehen. Für die Analyse wurden die Vorhersagen des Modells von der Wortsequenz- auf die Dokumentebene aggregiert. Für die Entwicklung des Machine-Learning-Modells wurde auf einen Korpus von 15.176 Dokumenten annotiert. Die Annotation fand dabei auf Wortsequenz-Ebene statt, um eine hohe Interpretierbarkeit und Qualität des Datensatzes zu gewährleisten. Das Training der Modelle fand auf 80 Prozent der Dokumente im Korpus und die Evaluation auf den restlichen 20 Prozent der Daten statt. Über alle Modelle hinweg können dabei sehr gute Ergebnisse erzielt werden. Über alle Teilqualifikationen und Berufsbereiche hinweg (Micro-Scores) hat das Modell auf Dokument-Ebene im Durchschnitt eine Precision von .87, einen Recall von .85 und einen F1-Score von .86. Die Macro-Scores liegen mit einer Precision von .82 und einem Recall von .77 leicht niedriger, weisen aber darauf hin, dass das Modell auch in der Breite gut funktioniert.

Limitationen: Das Modell funktioniert in der Breite sehr gut, jedoch ist die Performanz auf seltenen Teilqualifikationen gemischt. Je häufiger eine Teilqualifikation vorkommt, desto mehr Beispiele konnten trainiert werden, was zu einer besseren statistischen Generalisierung der Konzepte führt. Eine weitere Limitation ist die Abhängigkeit von einem Berufsklassifikationsalgorithmus. Eine Voraussetzung für die gezielte Extraktion von Tätigkeitsfeldern ist die Beschränkung auf die Zielberufe, für die die Tätigkeitsfelder entwickelt wurden. In der Theorie ist diese Einschränkung losgelöst von der Leistungsfähigkeit des Modells. In der Praxis führt ein schlechtes Berufsklassifikationsmodell zu nachgelagerten schlechteren Ergebnissen der Extraktion von Teilqualifikationen. Daraus lässt sich ableiten, dass innerhalb von Bereichen, auf denen das Modell angewendet wurde, Interpretationen sinnvoll sind, die Generalisierbarkeit über alle Berufe eingeschränkt ist. Für die Zukunft muss insbesondere das KldB-Extraktions-Modell neu entwickelt werden.

2.4 Landkreise und kreisfreie Städte nach der NUTS-Klassifikation

Taxonomie: Die regionalen Daten werden mithilfe der Nomenclature of Territorial Units (NUTS) Klassifikation der europäischen Union in der Version vom 01.01.2021 bestimmt. Die NUTS-Klassifikation ist ein hierarchisches System, welches die sozioökonomischen Regionen der europäischen Union auf unterschiedlichen Ebenen definiert.

Die hierarchische Anordnung der NUTS-Klassifikation ermöglicht es Daten über Regionen hinweg zu aggregieren. Alle Landkreise in einem Bundesland können so zusammengefasst werden.

Tabelle 2 Überblick über die Ebenen der NUTS-Klassifikation

Stelle	Ebene	Deutschland
NUTS1	Sozioökonomische Großregionen	Bundesländer
NUTS2	Basisregionen für regionalpolitische Maßnahmen	In der Regel: Regierungsbezirke
NUTS3	Kleine Regionen für spezifische Diagnosen	Landkreise und kreisfreie Städte

Hinweis: Da die aktuelle NUTS-Klassifikation vom 01.01.2021 stammt, ist hier noch nicht der aktuelle Kreisgebietsstand (aktuell: 400 Landkreise und kreisfreie Städte) abgebildet. Die Stadt Eisenach wurde im Juni 2021 in den Wartburgkreis aufgenommen.

Anreicherung: Der Ort der Ausschreibung ist definiert über den Ort an dem die Arbeit erbracht werden soll. Der Ort (z.B. Berlin) wird von TextKernel extrahiert und mit der Geo-Position (Längen- und Breitengrad) angereichert. Die Geo-Position wird in der Datenbank der Stellenausschreibungen mithilfe von PostGIS² auf die NUTS-3-Ebene gemapped.

Hinweis: Für ca. 12% der Stellen kann kein eindeutiger Ort bestimmt werden (August 2022: 11,4%). Diese Daten werden in die Auswertungen im Jobmonitor auf den Report-Seiten einbezogen. Wenn auf den Seiten „Top 10“, „Zeitreihe“ oder „Vergleich“ der Filter Region für Deutschland eingestellt wird, werden ausschließlich Stellen einbezogen, für die eine Region extrahiert werden konnte.

3 Datenaufbereitung

3.1 Deduplizierung

Die Stellenanzeigen werden von X Jobportalen und Y Unternehmensseiten gesammelt. Manche Stellen werden auf unterschiedlichen Portalen veröffentlicht, weshalb Duplikate erkannt und für die Analyse entfernt werden müssen. Jede Stellenanzeige, die gesammelt wird, erhält eine eigene Posting-ID. Wenn unterschiedliche Stellen einen Job beschreiben, werden sie unter einer Job-ID zusammengefasst.

TextKernel benutzt einen Algorithmus, um Duplikate zu erkennen. Dieser bezieht unter anderem den Job-Titel, die Text-Ähnlichkeit, sowie unterschiedliche Meta-Daten in die Analyse ein. Mehr Informationen zur Duplikate-Erkennung finden Sie auf der Seite von TextKernel³.

Tabelle 3 Anzahl der Stellenausschreibungen und Jobs von 2018 bis 2021

Jahr	Stellenanzeigen (in Millionen)	Jobs (in Millionen)	Anteil Duplikate
2018	25,2	9,1	63,9%
2019	64,7	12,9	76,4%
2020	57,3	13	77,2%
2021	71,6	16	77,6%
2022	79,6	15,1	81,1%

² <http://postgis.net/>

³ <https://www.textkernel.com/newsroom/online-job-postings-have-many-duplicates-but-how-can-you-detect-them-if-they-are-not-exact-copies-of-each-other/>

Für den Jobmonitor werden Stellen nur einmal gezählt. Für das Veröffentlichungsdatum und den zugeordneten Beruf wird das Datum der ersten Ausschreibung genutzt. Die transversalen Kompetenzen werden jeweils über alle Stellen, die zu einem Job gehören, extrahiert und der Job-ID hinzugefügt. In Tabelle X ist der Verlauf von 2018 bis 2021 abgebildet.

3.2 Bereinigung von Stellenausschreibungen in der Zeitarbeit

Unternehmen der Arbeitnehmerüberlassung unterscheiden sich bei der Ausschreibung von Stellen deutlich von direkten Arbeitgebern (vgl. Burstedde et al., 2020). Sie schreiben ihre Stellen nicht nur viel häufiger aus, sondern streuen diese auch über mehr Portale. Darüber hinaus ist die Suche nach Fachkräften auch Teil ihrer Portfoliobildung. Stellen werden über den unmittelbaren Bedarf hinaus ausgeschrieben, um Arbeitskräfte für künftige Kundenbedarfe vorhalten zu können. Das führt zu einem vergleichsweise hohen Anteil an Stellenausschreibungen vonseiten der Arbeitnehmerüberlassung. Bleibt dieser Umstand bei der Analyse von Stellenanzeigen unberücksichtigt, werden die von Zeitarbeitsunternehmen nachgefragten Berufe überschätzt. Würden die Stellen allerdings vollständig aus der Analyse herausgenommen, würden die Berufe in der Zeitarbeit wiederum unterschätzt. Denn zumindest in Teilen handelt sich hierbei um Stellenanzeigen, hinter denen auch echte Jobs stehen. Derzeit arbeiten etwa 800.000 sozialversicherungspflichtig Beschäftigte in der Zeit- bzw. Leiharbeit (vgl. Statistik der Bundesagentur für Arbeit 2023).

Um den Fachkräftebedarf in der Arbeitnehmerüberlassung weder zu unterschätzen noch zu überschätzen, arbeiten das Institut der deutschen Wirtschaft und auch das Kompetenzzentrum Fachkräftesicherung (KoFa) in entsprechenden Studien mit einem Zeitarbeitsgewicht. Das Zeitarbeitsgewicht wird dadurch bestimmt, dass die gemeldeten Stellen bei der Bundesagentur für Arbeit (BA) mit Hilfe der in der IAB-Stellenerhebung ermittelten Meldequoten auf gemeldete, offene Stellen hochgerechnet werden. Ohne die Stellen aus der Zeitarbeit ergäbe sich besagte Unterschätzung der offenen Stellen, mit den Stellen der Zeitarbeit eine Überschätzung der offenen, gemeldeten Stellen (vgl. Burstedde et al. 2020). Der Anteil, mit dem die offenen, gemeldeten Stellen hochgerechnet werden müssen, um möglichst nah an die tatsächlich offenen Stellen laut IAB-Stellenerhebung zu kommen, wird als Zeitarbeitsgewicht bezeichnet. Es liegt sowohl der vorliegenden Studie zugrunde als fortan auch dem Jobmonitor. Analog zur Methodik des IW Köln wurde mit einem rollierenden Drei-Jahres-Durchschnitt gerechnet. Im Gegensatz zu den bereits vollständig erfassten Jahren bis einschließlich 2022 wurde für das laufende Jahr 2023 mit einem rollierenden Durchschnitt (über die vergangenen drei Jahre) gerechnet.

Es gibt zwei Möglichkeiten, Stellen aus der Arbeitnehmerüberlassung identifizieren zu können. Der erste Weg führt über die Identifizierung des Wirtschaftszweiges (WZ78.2x WZ78.3x). Informationen zum Wirtschaftszweig liegen allerdings nur für Stellenanzeigen vor, die einem Unternehmen zugeordnet werden können, für das es auch Brancheninformationen gibt. Bei etwa 25 Prozent der Stellenanzeigen ist das nicht der Fall, wodurch Stellen aus der Arbeitnehmerüberlassung unterschätzt würden. Der zweite Weg führt über den Indikator „advertiser_type“, mit dem Textkernel zwischen direkten Arbeitgeber:innen und Personalagenturen bzw. Personalvermittlungen unterscheidet.

Ein Vergleich beider Wege ergab Folgendes: Über die Hälfte der Stellenanzeigen, die nicht aus den Wirtschaftszweigen 78.2x und 78.3x stammten, aber mit „Personalvermittlung“ gekennzeichnet waren, waren der Arbeitnehmerüberlassung zuzuordnen. Wir gehen weiterhin davon aus, dass Stellen der Personalvermittlung, die nicht den WZ 78.2x/3x zugeordnet sind und bei denen es sich nicht um Stellen der Arbeitnehmerüberlassung handelt, schwerer durch die eingesetzte Deduplizierung erfasst werden können. Hintergrund ist, dass Personalvermittler ihre Klienten bewusst nicht preisgeben möchten, Stellengesuche anders formulieren als Unternehmen der ANÜ und Gesuche auch nicht zeitgleich veröffentlichen. Die vorliegende Studie nutzt schließlich den Indikator „advertiser_type“. Er ist die bessere Alternative, um Stellen der ANÜ zu identifizieren (nähere Informationen zum Indikator unter: <https://www.textkernel.com/de/jobfeed-de/deep-learning-in-jobfeed/>). Um künftig den Einfluss von Zeitarbeitsstellen – sowie anderer potenziell bedarfsverzerrender Faktoren – noch genauer berücksichtigen zu können, werden die entsprechenden Analysemethoden weiter verfeinert.

Filterung v1 (ab 2023)

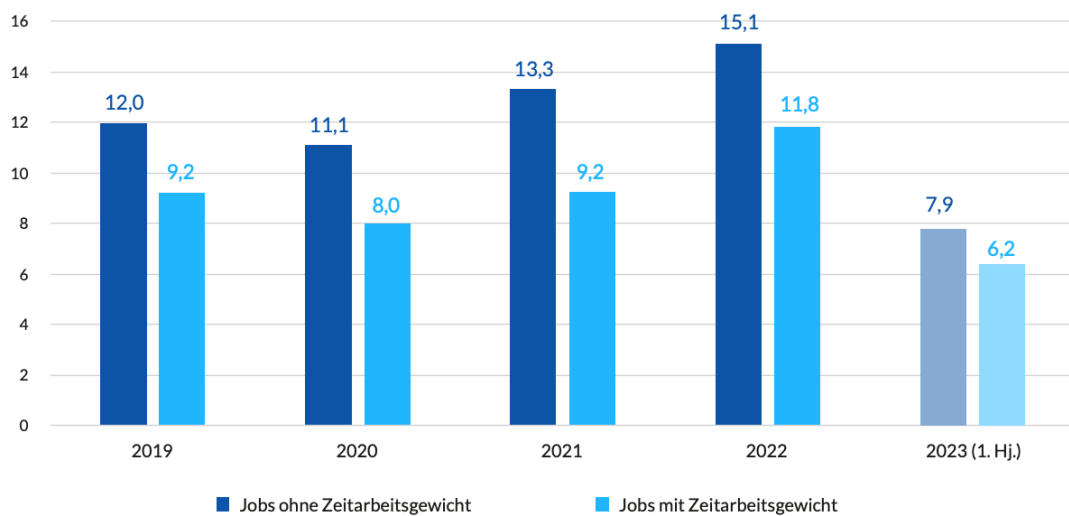
Tabelle 4 Jahresspezifische Quote für die tatsächliche Anzahl von neuen Stellen in der Zeitarbeit

Jahr	Quote
2018	0,454
2019	0,494
2020	0,383
2021	0,374
2022	0,481
2023	0,481

Für die Filterung benutzen wir jahresspezifische Quoten und ein gewichtetes Sampling der Stellenanzeigen. Stellen die nicht als Personalvermittlung gekennzeichnet erhalten ein Stichprobengewicht von 1 und werden alle in den Datensatz übernommen. Für Stellen in der Personalvermittlung wird ein Stichprobengewicht äquivalent zu der jahresspezifischen Quote angenommen.

$$\text{Einbeziehung}_{\text{Posting}} = \begin{cases} 1; & \text{wenn nicht Personalvermittlung} \\ 1 - \text{Gewicht}_{\text{Jahr}}; & \text{wenn Personalvermittlung} \end{cases}$$

Anzahl Online-Stellenanzeigen (Jobs) im Jobmonitor (Zugänge pro Jahr) in Millionen



Quelle: Eigene Berechnung.

JOBMONITOR | BertelsmannStiftung

Abbildung 4 Anzahl der Jobs im Jobmonitor vor und nach der Bereinigung (Herdin et al. 2023)

Filterung v2 (ab Juli 2026)

Eine Herausforderung bei der Berechnung der Quote, ist dass sie erst mit Verzögerung aktualisiert wird und an dem Zeitpunkt, an dem die Daten für den Jobmonitor berechnet werden, jeweils noch nicht verfügbar sind. Eine jahresspezifische Analyse ergab, dass das 10-Jahres-Mittel bei ungefähr bei 50% liegt. Deshalb wird ab Juli 2026 und rückwirkend für alle Jahre ein gleichmäßiges Gewicht von 50% angenommen.

3.3 Aggregation

Die Daten werden einmal pro Nacht aggregiert. Für den Jobmonitor werden dabei monatliche Stellenausschreibungen (dedupliziert) gezählt und auf der NUTS-3-Ebene aggregiert. Weiterhin werden die Daten für die drei weiteren Filter KLDB-5-Ebene, die Teilqualifikationen und die transversalen Kompetenzen aggregiert.

Eine Neuerung mit Version 2 der Filterung ist, dass Praktikumsstellen ausgeschlossen werden. Für die Extraktion von Anstellungsarten wird ein neues Modell von Baskaran (2026) verwendet.

4 Metriken

Auf dem Jobmonitor werden unterschiedliche Metriken berechnet. Im Folgenden werden die wichtigsten Berechnungen aufgeführt. Die Metriken können dabei für Filter aus dem Monat *month*, einem Beruf *kldb*, einer transversale Kompetenz *tsc* und einer Region *nuts* berechnet werden. Gezählt werden immer die Anzahl der Stellen *numberOfJobs*.

Hinweis: Alle Metriken lassen sich auch ohne die Angabe einer Region berechnen. Dabei ist zu beachten, dass die Region Deutschland (*nuts* = "DE") nicht zu den gleichen Ergebnissen führt wie eine Analyse ohne den Regions-Filter. Dies liegt daran, dass, wie in Kapitel 2 beschrieben, auch Stellen ohne eine Regionsangabe in der Datenbank abgelegt werden und diese auch nicht über den NUTS-Code aggregiert werden können.

4.1 Berufe

Für die Berufe ist die relevanteste Metrik die Anzahl der Stellenausschreibungen für einen Beruf für einen Monat.

$$totalJobs = \sum numberOfJobs_{month,kldb,nuts}$$

Die Anzahl der Jobs für einen Beruf ist kein direkter Indikator dafür, dass der Beruf an Relevanz gewinnt, da insgesamt das Volumen der Stellenausschreibungen über Zeit anwächst. Die Relevanz eines Berufs in einer Region kann deshalb auch als Anteil der Stellen für den Beruf an allen Berufen in der Region berechnet werden.

$$percentageTotalJobs = \frac{\sum numberOfJobs_{month,kldb,nuts}}{\sum numberOfJobs_{month,nuts}}$$

Der Vergleich zwischen Regionen wird weiterhin stark beeinflusst von der Bevölkerungszahl und der allgemeinen sozio-ökonomischen Struktur. Ein Ansatz für diese Unterschiede zu korrigieren, ist die Zahl der Stellenausschreibungen durch die Anzahl der Beschäftigten am Arbeitsort zu teilen. Die Metrik Stellen pro 10 000 Beschäftigte am Arbeitsort wird folgendermaßen berechnet

$$totalJobsPer10k = \frac{\sum numberOfJobs_{month,kldb,nuts}}{numberOfWorkers_{nuts}} * 10000$$

Die Zahlen zu den Beschäftigten am Arbeitsort werden jeweils nur unregelmäßig für ganz Deutschland auf Kreisebene von der Bundesagentur für Arbeit veröffentlicht. Deshalb werden für den Jobmonitor (Stand September 2022) die Zahlen zum Stichtag 31.12.2021 genutzt. (Quelle einfügen).

Die Veränderung der Stellen werden jeweils zum Vormonat und zum Vorjahresmonat berechnet – jeweils als absolute Differenz und prozentuale Differenz.

$$\begin{aligned} totalDifferenceJobsLastMonth \\ = \sum numberOfJobs_{month,kldb,nuts} - \sum numberOfJobs_{month-1,kldb,nuts} \end{aligned}$$

$$\begin{aligned} totalDifferenceJobsLastYear \\ = \sum numberOfJobs_{month,kldb,nuts} - \sum numberOfJobs_{month-12,kldb,nuts} \end{aligned}$$

$$percDiffTotalJobsLastMonth = \frac{\sum numberOfJobs_{month,kldb,nuts}}{\sum numberOfJobs_{month-1,kldb,nuts}} - 1$$

$$percDiffTotalJobsLastYear = \frac{\sum numberOfJobs_{month,kldb,nuts}}{\sum numberOfJobs_{month-12,kldb,nuts}} - 1$$

4.2 Soft Skills (transversale Kompetenzen)

Die relevante Metrik für die Soft Skills ist der Anteil der Stellen, in welchen ein bestimmter Skill gesucht wird. Dafür wird die Anzahl der Stellen mit dem Skill `numberOfJobsWithSkill` durch die gesamte Anzahl der Stellen für den Filter geteilt.

$$percentageTotalJobs_{TSC} = \frac{\sum numberOfJobsWithSkill_{month,tsc,nuts}}{\sum numberOfJobs_{month,nuts}}$$

Die Veränderung für die Soft Skills wird jeweils als Veränderung in Prozentpunkten berechnet.

$$\begin{aligned} percentageDiffJobsLastMonth_{tsc} \\ = percentageTotalJobs_{month,tsc,nuts} - percentageTotalJobs_{month-1,tsc,nuts} \end{aligned}$$

$$\begin{aligned} &percentageDiffJobsLastYear_{tsc} \\ &= percentageTotalJobs_{month,tsc,nuts} - percentageTotalJobs_{month-12,tsc,nuts} \end{aligned}$$

4.3 Teilqualifikationen

Die relevante Metrik für die Teilqualifikationen (TQ) ist der Anteil der Stellen, in welchen eine bestimmte Teilqualifikation gesucht wird. Der Nenner ist dabei die Anzahl der gesamten Stellen für einen Beruf. Dafür wird die Anzahl der Stellen mit dem Skill numberOfJobsWithTQ durch die gesamte Anzahl der Stellen für den Filter geteilt.

$$percentageTotalJobs_{TSC} = \frac{\sum numberOfJobsWithTQ_{month,tq,nuts,kldb}}{\sum numberOfJobs_{month,nuts,kldb}}$$

Die Veränderung für die Teilqualifikationen wird jeweils als Veränderung in Prozentpunkten berechnet.

$$\begin{aligned} &percentageDiffJobsLastMonth_{tq,kldb} \\ &= percentageTotalTQ_{month,tq,nuts,kldb} \\ &- percentageTotalJobs_{month-1,tq,nuts,kldb} \end{aligned}$$

$$\begin{aligned} &percentageDiffJobsLastYear_{tq,kldb} \\ &= percentageTotalJobs_{month,tq,nuts,kldb} \\ &- percentageTotalJobs_{month-12,tq,nuts,kldb} \end{aligned}$$

5 Verweise

Burstedde, A., Flake, R., Jansen, A., Malin, L., Risius, P., Seyda, S., Schirner, S., Werner, D. (2023): Die Messung des Fachkräftemangels. IW Report 59/2020. Köln, 21.11.2020.

https://www.iwkoeln.de/fileadmin/user_upload/Studien/Report/PDF/2020/IW-Report_2020_Messung_des_Fachkraeftemangels.pdf

Baskaran, R., Müller, J. (2023): Algorithmische Normalisierung von Job Titeln auf die Klassifikation der Berufe 2010. https://www.and-effect.com/publications/2022-11-21_technical_report_kldb.pdf

Baskaran, R. (2026): Extraction of contract types from German OJA. https://www.and-effect.com/publications/2026-06-29_technical_report_contract_type_v1.pdf

Herdin, G., Baskaran, R., Fingerhut, J., Müller, J. (2023): Das große Berufsranking – Wie sich die Nachfrage nach Berufen in Deutschland verändert. <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/das-grosse-berufe-ranking>

Müller, J. (2022, TQ): Machbarkeitsstudie: Teilqualifikationen in Online-Job-Anzeigen (OJA) – Methodenbericht zur automatisierten Extraktion von Teilqualifikationen für fünf Ausbildungsberufe. https://www.and-effect.com/publications/2022-05-13_methodenbericht_teilqualifikationen_v3.pdf

Müller, Johannes, (2023). Algorithmische Extraktion von Teilqualifikationen aus deutschsprachigen Online-Jobanzeigen. https://www.and-effect.com/publications/2023-03-16_methodenbericht_teilqualifikationen_v1.pdf

Müller, J., Fingerhut, J., Noack, M. (2022): Algorithmische Extraktion transversaler Kompetenzen nach ESCO v1.1 aus deutschsprachigen Online-Jobanzeigen – Methodenbericht. https://www.and-effect.com/publications/2022-09-25_methodenbericht_transversale_kompetenzen_v1.pdf

Müller, J. (2026): Classification of German job titles in online job postings using the KldB-2010 taxonomy – Model version 3. https://www.and-effect.com/publications/2026-06-29_technical_report_kldb_v3.pdf

Statistik der Bundesagentur für Arbeit (2023): Berichte: Blickpunkt Arbeitsmarkt – Entwicklungen in der Zeitarbeit, Nürnberg, Juli 2023. <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Statistiken/Themen-im-Fokus/Zeitarbeit/generische-Publikation/Arbeitsmarkt-Deutschland-Zeitarbeit-Aktuelle-Entwicklung.pdf>

Vicari, Basha (2014): Grad der standardisierten Zertifizierung des Berufs – Ein Indikator zur Messung institutioneller Eigenschaften von Berufen. Institut für Arbeitsmarkt- und Berufsforschung. https://doku.iab.de/fdz/reporte/2014/MR_04-14.pdf

6 Tabellenverzeichnis

Abbildung 1 Aggregation und Anreicherung der Online-Stellenanzeigen.	...2
Abbildung 2 Anzahl der Jobs im Jobmonitor vor und nach der Bereinigung.	...12
Tabelle 1 Überblick Klassifikation der Berufe 2010	...4
Tabelle 2 Überblick über die Ebenen der NUTS-Klassifikation	...6
Tabelle 3 Anzahl der Stellenausschreibungen und Jobs von 2018 bis 2021	...9
Tabelle 4 Gewichte für Stellen der Arbeitnehmerüberlassung nach Jahren	...10