



IW-Gutachten

Expertise eines methodischen Ansatzes zur Identifizierung von beruflichen Übergangspfaden in der Automobil- und Zulieferindustrie in Baden-Württemberg

Regina Flake / Henry Goecke / Helen Hickmann / Armin Mertens / Sebastian Schirner /
Susanne Seyda

Auftraggeber: Bertelsmann Stiftung
Köln, 16.12.2020

Inhaltsverzeichnis

Zusammenfassung	3
1 Einleitung	4
2 Identifikation von relevanten Berufen in der Automobil- und Zulieferindustrie	5
2.1 Klassifikation der Berufe 2010 und Beschäftigtendaten nach KldB	5
2.2 Klassifikationen der Berufe nach ISCO	9
2.3 Gruppierung der Berufe anhand steigender und sinkender Nachfrage	11
3 Analyse der beruflichen Kompetenzen in der Automobil- und Zulieferindustrie	14
3.1 Analyse von Berufsinformationen	14
3.1.1 Berufenet, Aus- und Fortbildungsordnungen sowie Studienordnungen	14
3.1.2 ESCO	15
3.1.3 Weitere Datenquellen	16
3.2 Analyse von Stellenausschreibungen der Automobil- und Zulieferindustrie	16
3.2.1 Identifikation von relevanten Stellenausschreibungen der Automobil- und Zulieferindustrie	18
3.2.2 Identifikation von Kompetenzen in Stellenausschreibungen der Automobil- und Zulieferindustrie	27
3.3 Zusammenführung der Arbeiten	31
4 Ähnlichkeitsanalyse zwischen Berufen	32
5 Optionsmodul: Änderungen innerhalb von Berufen	36
6 Fazit	37
7 Literatur	38
Tabellenverzeichnis	43
Abbildungsverzeichnis	44

JEL-Klassifikation:

C45 – Neural Networks and Related Topics

J32 – Labor Demand

J24 – Human Capital; Skills; Occupational Choice; Labor Productivity

Zusammenfassung

Vor dem Hintergrund laufender Transformationsprozesse, ausgelöst durch immer weitere technologische Innovationen, verändern sich Berufsbilder und damit auch die geforderten Kompetenzprofile von Beschäftigten. Dadurch, dass bestimmte Berufe vermehrt und andere Berufe hingegen weniger nachgefragt werden, steht in besonderem Maße die technologieintensive Automobil- und Zulieferindustrie vor der Herausforderung, den Transformationsprozess auch auf personeller Ebene zu gestalten und so dem (drohenden) Fachkräftemangel entgegenzuwirken. Dies kann durch entsprechende Qualifizierungen erreicht werden, die es Beschäftigten ermöglichen, ihre Kompetenzprofile so zu erweitern, dass sie neue Aufgaben übernehmen und in vermehrt nachgefragte oder auch neu entstandene Berufe wechseln können, wenn ursprüngliche Tätigkeiten entfallen.

Zur Bestimmung der Qualifizierungsbedarfe und des -potenzials können Analysen der veränderten Berufsnachfrage und der entsprechenden Kompetenzprofile dienen. Die drei folgenden Schritte führen zu einer empirischen Grundlage, von der ausgehend Übergangspfade entwickelt werden können: Zunächst sind die relevanten Berufe in der Automobil- und Zulieferindustrie zu identifizieren und in Berufe mit steigender und sinkender Nachfrage zu gruppieren. Als grundlegende Taxonomie bietet sich hierfür die Klassifikation der Berufe 2010 (KldB 2010) an und als Datengrundlage die Beschäftigtenstatistik der Bundesagentur für Arbeit (BA). Zweitens sind Kompetenzprofile zu erstellen. In diesem Arbeitsschritt empfehlen sich Big Data Methoden, um die Informationen aus verschiedenen Datenquellen, wie dem Berufenet der BA und aktuellen Stellenanzeigen, zusammenzuführen. Der dritte Schritt beinhaltet eine Ähnlichkeitsanalyse der Kompetenzprofile zwischen den identifizierten Berufen vor dem Hintergrund der wachsenden beziehungsweise sinkenden Nachfrage nach diesen Berufen. Hierfür werden Big Data Methoden angewandt und mit fundiertem Praxiswissen hinsichtlich der Einordnung und Bewertung der geforderten Kompetenzen kombiniert. Des Weiteren könnte in einem nächsten (optionalen) Schritt eine Analyse der Veränderungen innerhalb der Berufe durchgeführt werden, um konkrete Hinweise für die Gestaltung von Qualifizierungsangeboten zu erhalten.

1 Einleitung

Transformationsprozesse lösen Anpassungsbedarf bei den Kompetenzen der Beschäftigten aus. Dies kann auf der einen Seite dazu führen, dass bestimmte Berufe und die in ihnen verankerten Kompetenzen vermehrt nachgefragt werden, andere hingegen weniger. Eine solche Situation ist im Strukturwandel üblich, bedeutet aber in Zeiten bedeutender und rascher Technologieänderungen eine große Herausforderung für Unternehmen und Mitarbeiter. Im Idealfall gelingt es, Beschäftigte aus weniger nachgefragten Berufen so weiterzuqualifizieren, dass sie vermehrt Tätigkeiten in anderen Berufen, die eine wachsende Nachfrage verzeichnen, übernehmen können. Die Gestaltung solcher Übergänge setzt eine differenzierte Analyse der Kompetenzprofile für verschiedene Berufe voraus.

In der hier vorgeschlagenen Methodik soll in vier Schritten die Entwicklung von Übergangspfaden in der Automobil- und Zulieferindustrie vorbereitet werden:

1. Identifikation von relevanten Berufen: In Schritt 1 sollen relevante Berufe in der Automobil- und Zulieferindustrie identifiziert werden. Diese sollen über verschiedene Analysen, zum Beispiel zur Beschäftigtenentwicklung oder auch zur Fachkräfteengpasssituation, in Berufe mit steigender und sinkender Nachfrage gruppiert werden (Kapitel 2).
2. Erstellung von Kompetenzprofilen: In Schritt 2 werden für die identifizierten Berufe Kompetenzprofile erarbeitet. Hierfür werden zum einen Quellen ausgewertet, die aufzeigen, welche Kompetenzen in verschiedenen Berufsprofilen verankert sind (Kapitel 3.1). Da Tätigkeits- und Kompetenzbeschreibungen häufig technologieneutral formuliert sind, werden zum anderen aktuelle Stellenanzeigen ausgewertet, die die nachgefragten Kompetenzen aus Unternehmenssicht konkretisieren (Kapitel 3.2). Aus der Synthese ergeben sich berufsspezifische Kompetenzprofile (Kapitel 3.3).
3. Ähnlichkeitsanalyse zwischen Berufen: In Schritt 3 wird die Ähnlichkeit von Berufen basierend auf den Kompetenzprofilen analysiert (Kapitel 4). Dabei wird insbesondere überprüft, in welchen Kompetenzen sich Berufe mit wachsender beziehungsweise sinkender Nachfrage ähneln. Ein Ziel dieser Ähnlichkeitsanalyse ist es, Vorarbeiten für die Entwicklung von Übergangspfaden von weniger gefragten oder gefährdeten Berufen in wachsende Berufsgruppen zu entwickeln.
4. Änderungen innerhalb von Berufen (optional): Im vierten und letzten Schritt können ergänzend Änderungen innerhalb von Berufen analysiert werden. So können Qualifizierungspfade abgeleitet werden, die beispielsweise ältere Beschäftigte in einem Beruf in die Lage versetzen, aktuelle Anforderungen zu erfüllen.

2 Identifikation von relevanten Berufen in der Automobil- und Zulieferindustrie

In einem ersten Schritt sollen für die Automobil- und Zulieferindustrie relevante Berufe identifiziert werden. Hierfür werden im Folgenden verschiedene Taxonomien betrachtet und dargelegt, inwieweit sie sich für das geplante Forschungsprojekt eignen.

Für eine passgenaue Analyse ist es wichtig, eine Taxonomie zu finden, die das in Deutschland nach wie vor vorherrschende Berufsprinzip bestmöglich abbildet, da in Deutschland Berufsabschlüsse eine große Rolle für das Einstellungsverhalten von Arbeitgebern spielen. Dies lässt sich damit erklären, dass Unternehmen klare Erwartungen haben, welche Kompetenzen in den jeweiligen Berufen in Aus- und Fortbildung sowie Studium vermittelt werden.

2.1 Klassifikation der Berufe 2010 und Beschäftigtendaten nach KldB

Die zentrale Klassifikation für den deutschen Arbeitsmarkt ist die Klassifikation der Berufe 2010 (KldB 2010) der Bundesagentur für Arbeit (BA). Die KldB 2010 berücksichtigt die Besonderheiten des deutschen Arbeitsmarktes mit seiner berufsspezifischen Strukturierung. Darüber hinaus findet sich diese Klassifizierung in vielen Datensätzen als Strukturvariable wieder, die für die weiteren Analysen relevant sind.

Die KldB 2010 besteht aus fünf Gliederungsebenen. Die Gliederungsebenen werden immer feiner: Auf der ersten Ebene werden 10 Berufsbereiche, auf der zweiten Ebene 37 Berufshauptgruppen, auf der dritten Ebene 144 Berufsgruppen, auf der vierten Ebene 700 Berufsuntergruppen und auf der fünften und untersten Gliederungsebene (5-Steller) 1.286 Berufsgattungen unterschieden. In einer Berufsgattung sind in der Regel mehrere ähnliche Berufe zusammengefasst.¹ Die KldB gliedert Berufsgattungen anhand ihrer Berufsfachlichkeit, das heißt, anhand der Ähnlichkeit von Berufen innerhalb der Berufsgattung in Bezug auf die sie auszeichnenden Tätigkeiten, Kenntnisse und Fertigkeiten sowie anhand ihres Anforderungsniveaus, also der Komplexität der auszuübenden Tätigkeit (Tabelle 2-1). Die KldB 2010 unterscheidet zudem zwischen vier Anforderungsniveaus, die mit bestimmten Bildungsabschlüssen verbunden sind: Helfer (An- und Ungelernte), Fachkräfte (abgeschlossene Berufsausbildung), Spezialisten (Aufstiegsfortbildung, wie Meister, Techniker oder Fachwirt, sowie Bachelor) und Experten (Master oder Diplom sowie Promotion) (Tabelle 2-1).

¹ Trotz dieses Unterschieds werden die Begriffe „Berufsgattung“ und „Beruf“ in den folgenden Kapiteln synonym verwendet. Es ist dennoch wichtig, diese Differenzierung bei der Interpretation der Analysen zu berücksichtigen. So ist denkbar, dass sich die Berufe in unterschiedliche Richtungen entwickeln und die Berufsgattung als Ganzes diese Entwicklungen nur teilweise abbildet. Dies könnte im Stakeholder Prozess adressiert werden.

Tabelle 2-1: Anforderungsniveaus in der KldB 2010

Anforderungs-niveau	Bezeichnung	Typische Qualifikation
1	Helfer	Geringqualifiziert
2	Fachkraft	Berufsausbildung (mindestens zweijährig)
3	Spezialist	Fortbildungsabschluss (z. B. Meister, Techniker, Fachwirt) oder Bachelor mit wenig Berufserfahrung
4	Experte	Diplom, Master oder Bachelor mit viel Berufserfahrung

Quelle: IW-Darstellung auf Basis von BA (2011a)

Die KldB 2010 ist im Rahmen des Meldeverfahrens zur Sozialversicherung Grundlage für die Verschlüsselung der ausgeübten Tätigkeit, die der Arbeitgeber anzugeben hat, und fließt dementsprechend in die Beschäftigtenstatistik der BA ein – ein Datensatz, der für die weitere Analyse relevant ist (BA, 2019).² Hierfür gibt es ein Schlüsselverzeichnis, welches die Arbeitgeber nutzen können. Entscheidend für die Arbeitgebermeldung ist dabei allein die Tätigkeit, die der Beschäftigte im Betrieb ausübt, nicht der erlernte Beruf. Wenn mehrere Tätigkeitsbezeichnungen zu einer Stelle passen, wird der Beschäftigte dem Beruf zugeordnet, aus dem er überwiegend Tätigkeiten ausübt. In das Schlüsselverzeichniss der BA werden regelmäßig neue Berufsbezeichnungen aufgenommen und entsprechend der KldB 2010 zugeordnet. Diese Anpassungen leiten sich aus dem Wandel der Berufslandschaft ab (Entstehung neuer Berufe, Wegfall anderer Berufe, Veränderungen der Bezeichnungen etc.). Die KldB 2010 verändert sich hierdurch nicht. Es kann daher sein, dass sich Tätigkeiten innerhalb der Berufe ändern oder immer mehr Tätigkeiten aus anderen Berufen mit übernommen werden („hybride Berufe“). Welche Kompetenzen in dem jeweiligen Beruf aktuell von Bedeutung sind, wird über die Analyse der Kompetenzen auf Basis von Berufsinformationen (Kapitel 3) sowie der Stellenausschreibungen (Kapitel 3.2) identifiziert. Veränderungen innerhalb von Berufen könnten über Zeitvergleiche analysiert werden (Kapitel 5). Auch im Rahmen der BA-Arbeitslosen- und Stellenstatistik, die ebenfalls für die in dieser Expertise vorgestellte Methodik relevant ist, wird die gewünschte berufliche Tätigkeit der zu besetzenden Arbeitsstelle mit der KldB 2010 verschlüsselt.

Arbeits- und Ausbildungsmarktdaten auf Ebene der KldB 2010 werden von der BA bereitgestellt. Sie stehen teilweise kostenfrei auf der Seite der BA zur Verfügung. Beruflich und regional gegliederte Daten (zum Beispiel auf Ebene der Bundesländer) können als Sonderauswertungen kostenpflichtig bei der BA bestellt werden. Das Institut der deutschen Wirtschaft (IW) führt regelmäßig Daten von der BA zu sozialversicherungspflichtig Beschäftigten, zu gemeldeten offenen Stellen und zu Arbeitslosen mit Daten zum Ausbildungsmarkt auf Ebene der KldB-5-Steller zusammen und nutzt diese so erstellte Datenbank für Analysen zur Fachkräftesituation auf dem Arbeits- und Ausbildungsmarkt (siehe Burstedde et al., 2020).

² Die Meldepflicht ist im Sozialgesetzbuch (SGB) Viertes Buch (IV) § 28a geregelt (http://www.gesetze-im-internet.de/sgb_4/_28a.html).

Die Daten zum Bestand der sozialversicherungspflichtig Beschäftigten sind auf Ebene der Berufsgattungen unter anderem differenziert nach Geschlecht, Altersgruppe, Herkunft (Deutschland, Europäische Union, Drittstaat), Arbeitsumfang (Vollzeit oder Teilzeit), Betriebsgrößenklassen und Wirtschaftszweigen nach der Klassifikation der Wirtschaftszweige (WZ 2008) verfügbar. Auch diese Daten sind im Rahmen von Sonderauswertungen bei der BA erhältlich und liegen aufgegliedert nach der KldB 2010 seit einschließlich 2013 vor (BA, 2020).

Über die KldB 2010 lassen sich relevante Berufe für die Automobil- und Zulieferindustrie identifizieren. So enthält zum Beispiel die Berufsgruppe 252 „Fahrzeug-, Luft-, Raumfahrt- und Schiffbautechnik“ mit der Untergruppe 2521 „Berufe in der Kraftfahrzeugtechnik“ viele relevante Berufsgruppen. Da in der Automobilbranche jedoch auch andere Berufe relevant sind, die nicht spezifisch für diese Industrie sind, ist zusätzlich eine Branchenanalyse notwendig, um alle relevanten Berufe zu identifizieren. Es existieren zahlreiche Berufe, die in vielen Branchen relevant sind. So werden beispielsweise viele technische Berufe zum Aufbau und zur Wartung von Produktionsanlagen in einer Vielzahl von Branchen benötigt. Um Beschäftigtendaten der BA auf Ebene der KldB-5-Steller nach Branche untersuchen zu können, kann auf die WZ 2008, die Klassifikation der Wirtschaftszweige, zurückgegriffen werden. Die Zuordnung der Beschäftigten zu den Wirtschaftszweigen erfolgt über die Betriebsnummer, die von der BA gemeinsam mit dem Unternehmen vergeben wird. Ist ein Unternehmen in verschiedenen Niederlassungen in verschiedenen Wirtschaftszweigen tätig, erhält es verschiedene Betriebsnummern und die Beschäftigten werden den verschiedenen Wirtschaftszweigen zugeordnet. Ist ein Unternehmen in verschiedenen Wirtschaftszweigen an einem Standort tätig, wird nur eine Betriebsnummer und ein Wirtschaftszweig zugeordnet (BA, 2017).

Abbildung 2-1: Die Automobil- und Zulieferindustrie in der Klassifikation der Wirtschaftszweige (WZ 2008)

Abteilung 29 „Herstellung von Kraftwagen und Kraftwagenteilen“

29	Herstellung von Kraftwagen und Kraftwagenteilen	
29.1	Herstellung von Kraftwagen und Kraftwagenmotoren	
29.10	Herstellung von Kraftwagen und Kraftwagenmotoren	2910
29.10.1	Herstellung von Personenkraftwagen und Personenkraftwagenmotoren	
29.10.2	Herstellung von Nutzkraftwagen und Nutzkraftwagenmotoren	
29.2	Herstellung von Karosserien, Aufbauten und Anhängern	
29.20	Herstellung von Karosserien, Aufbauten und Anhängern	2920
29.20.0	Herstellung von Karosserien, Aufbauten und Anhängern	
29.3	Herstellung von Teilen und Zubehör für Kraftwagen	
29.31	Herstellung elektrischer und elektronischer Ausrüstungsgegenstände für Kraftwagen	2930*
29.31.0	Herstellung elektrischer und elektronischer Ausrüstungsgegenstände für Kraftwagen	
29.32	Herstellung von sonstigen Teilen und sonstigem Zubehör für Kraftwagen	2930*
29.32.0	Herstellung von sonstigen Teilen und sonstigem Zubehör für Kraftwagen	

Quelle: Statistisches Bundesamt, 2018

Abteilung 29 der WZ 2008 umfasst die Herstellung von Kraftwagen und Kraftwagenteilen. In dieser Branche ist ein Großteil der für die Analyse relevanten Berufsgattungen zu finden. Sie umfasst nicht nur die eigentliche Herstellung von Kraftwagen und Kraftwagenmotoren, sondern explizit auch die Herstellung von Ausrüstungsgegenständen und Zubehör für Kraftwagen und damit Automobilzulieferer. Abbildung 2-1 bietet einen Überblick über die Abteilung 29. Nicht enthalten sind die Herstellung von Elektromotoren, von Kolben und Kolbenringen, von Vergasern sowie von Beleuchtungseinrichtungen für Kraftfahrzeuge, die in der Abteilung 27 zu finden sind. Wie diese Beispiele zeigen, können auch andere Abteilungen neben der Abteilung 29 relevant für die Abgrenzung der Berufsgattungen sein. Für die finale Abgrenzung der für das Projekt relevanten Berufsgattungen sollten deshalb auch weitere Abteilungen geprüft und in Abstimmung mit dem Auftraggeber und den Stakeholdern gegebenenfalls einbezogen werden.

In unmittelbarer Nachbarschaft der Abteilung 29 befinden sich die Abteilungen 28 „Maschinenbau“ und 30 „Sonstiger Fahrzeugbau“. Beide Kategorien bilden allerdings weniger spezifisch die Berufsgattungen der (Kraft-)Fahrzeugindustrie ab. So ist die Berufsgattung mit den meisten Beschäftigten in Abteilung 30 beispielsweise die Fachkraft für Luft- und Raumfahrttechnik, gefolgt von den Experten für Luft- und Raumfahrt. In Abteilung 28 sind die Fachkräfte für Maschinenbau- und Betriebstechnik und spanende Metallbearbeitung die beiden Berufsgattungen mit den meisten Beschäftigten.

Um die relevanten Berufe für die Automobil- und Zulieferindustrie zu ermitteln, wird die Anzahl der sozialversicherungspflichtig Beschäftigten nach KldB 2010 und nach WZ 2008 gemeinsam betrachtet. Mit dieser Aufstellung lässt sich erkennen, welche Berufe für die Branche quantitativ relevant sind (Relevanz). Die Daten für alle 88 Wirtschaftszweige und 1.286 Berufsgattungen können als kostenpflichtige Sonderauswertung sowohl auf Bundesebene als auch für Bundesländer bei der BA bestellt werden. Die Abgrenzung kann damit spezifisch für Baden-Württemberg erfolgen.³ Die Berufsgattung mit der bundesweit größten Zahl sozialversicherungspflichtig Beschäftigter in Abteilung 29 ist die Fachkraft für Fahrzeugtechnik, gefolgt von den Experten der technischen Forschung und Entwicklung. Zur Fachkraft für Fahrzeugtechnik gehören die Ausbildungsberufe Karosserie- und Fahrzeugbaumechaniker/-in (Fachrichtungen Karosserie- und Fahrzeugbautechnik sowie Karosserieinstandhaltungstechnik) sowie Kraftfahrzeugmechatroniker/-in mit verschiedenen Schwerpunkten.

Neben der quantitativen Relevanz der Berufsgattungen lässt sich prüfen, ob eine Berufsgattung nur für eine Branche relevant ist oder auch für andere Branchen (Spezifität). Hier zeigt sich beispielsweise, dass 59 Prozent der Fachkräfte der Fahrzeugtechnik im Bereich Handel mit Kraftfahrzeugen; Instandhaltung und Reparatur von Kraftfahrzeugen (Abteilung 45) beschäftigt sind. 20 Prozent aller beschäftigten Fachkräfte der Fahrzeugtechnik werden Abteilung 29 zugeordnet. Die verbleibenden rund 20 Prozent der Beschäftigten verteilen sich mit kleinen Beschäftigungsanteilen auf beinahe alle anderen Wirtschaftszweige.

³ Im weiteren Verlauf des Dokuments beziehen sich die Beispiele – sofern nicht anders beschrieben – auf die Daten auf Bundesebene.

Für die Abgrenzung der für das Projekt relevanten Automobilberufe muss ein Grenzwert für die Beschäftigung nach Wirtschaftszweig bestimmt werden, unterhalb dessen Berufsgattungen nicht mehr berücksichtigt werden. Dabei ist zu beachten, dass sich die Beschäftigung nicht gleichmäßig auf die vier Anforderungsniveaus verteilt. In Abteilung 29 sind die meisten Beschäftigten beruflich qualifiziert, gefolgt von deutlich weniger Experten und Spezialisten. Die Grenzwerte für die Relevanz der Berufe könnten daher innerhalb der Abteilung je Anforderungsniveau spezifisch festgelegt werden. So wäre sichergestellt, dass aus allen Anforderungsniveaus ausreichend Berufe berücksichtigt werden. Diese Grenzwerte können in absoluten Zahlen oder prozentual gesetzt werden. Ein Grenzwert von 1.000 Beschäftigten für alle Qualifikationsniveaus ergäbe beispielsweise 116 Berufsgattungen in Abteilung 29, die 93,3 Prozent aller Beschäftigten dieser Abteilung abdecken. Ein Grenzwert von 100 Beschäftigten würde zu 264 Berufsgattungen führen, die 99,0 Prozent der Beschäftigten umfassen. Grenzwerte von 1.000 Beschäftigten für Fachkräfte und jeweils 500 für Spezialisten und Experten würden zu 122 Berufsgattungen führen.

Ergänzend sollte zur weiteren Eingrenzung der relevanten Berufe mit Hilfe eines zweiten Grenzwerts die Spezifität der Berufsgattungen berücksichtigt werden. Es könnte zum Beispiel festgelegt werden, dass mindestens 15 Prozent der Beschäftigten einer Berufsgattung in der Automobil- und Zulieferindustrie tätig sein müssen. Damit würde sichergestellt, dass nur wenig branchenspezifische Berufe wie Fachkräfte für Büro- und Sekretariat oder für die Reinigung nicht als – im Sinne der hier vorgelegten Expertise – relevante Berufe identifiziert werden. Der Grenzwert würde die oben genannten 116 Berufe (bei einem Grenzwert von 1.000 Beschäftigten) auf 30 Berufsgattungen reduzieren. Aufgrund der hohen Zahl an Wirtschaftszweigen und der Verbreitung vieler technischer Berufsgattungen in anderen Wirtschaftszweigen sollte der Grenzwert der Spezifität nicht zu hoch gewählt werden.

Die Liste der als relevant identifizierten Berufe sollte im Projektverlauf mit den Stakeholdern abgestimmt werden. So wird zum einen sichergestellt, dass in der weiteren empirischen Analyse keine Berufe im Fokus stehen, die für die Projektpartner nicht von Interesse sind, und zum anderen können so gegebenenfalls noch Berufe aufgenommen werden, die zwar eine quantitativ geringe Bedeutung haben, aber aus anderen Gründen von Interesse sind. So können auch qualitative Kriterien bei der Berufsauswahl berücksichtigt werden.

2.2 Klassifikationen der Berufe nach ISCO

Neben der KldB 2010 als nationale Klassifikation gibt es auch internationale Klassifikationen von Berufen wie die International Standard Classification of Occupations (ISCO-08), die von der International Labour Organization (ILO) zur Verfügung gestellt wird. Wie Tabelle 2-2 zeigt, ist diese Klassifikation jedoch viel gröber als die KldB 2010, so dass sie nicht zu der in Deutschland typischen Berufestruktur passt.

Tabelle 2-2: Vergleich KldB 2010 und ISCO-08

Gliederungsebenen und Zahl der Kategorien

	ISCO-08	KldB2010
1. Ebene	10	10
2. Ebene	42	37
3. Ebene	128	144
4. Ebene	436	700
5. Ebene	-	1.286

Quelle: BA, 2011a

Dies lässt sich an der Berufsgattung des Experten für Kraftfahrzeugtechnik deutlich machen. Während jener in der KldB 2010 ein eigener 5-Steller ist (25214), wird er in ISCO-08 der Gruppe der Maschinenbauingenieure (2144) zugeordnet (Tabelle 2-3). Diese Gruppe ist jedoch sehr heterogen und enthält beispielsweise auch Experten aus dem Bereich der Papierverarbeitung oder Kältetechnik, die keine oder nur eine untergeordnete Bedeutung für die Automobil- und Zulieferindustrie haben.

Tabelle 2-3: Maschinenbauingenieure in der ISCO-08 und der KldB 2010

Umsteigeschlüssel

KldB 2010 (5-Steller)	Bezeichnungen der KldB2010 (5-Steller)	ISCO-08 (4-Steller)	Bezeichnungen der ISCO-08 (4-Steller)
23124	Berufe in der Papierverarbeitung und Verpackungstechnik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
24424	Berufe in der Schweiß- und Verbindungstechnik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
24514	Berufe in der Feinwerktechnik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
24524	Berufe in der Werkzeugtechnik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
25104	Berufe in der Maschinenbau- und Betriebstechnik (ohne Spezialisierung) - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
25134	Technische Servicekräfte in Wartung und Instandhaltung - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
25184	Berufe in der Maschinenbau- und Betriebstechnik (sonstige spezifische Tätigkeitsangabe) - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure

25214	Berufe in der Kraftfahrzeugtechnik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
25224	Berufe in der Land- und Baumaschinentechnik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
25234	Berufe in der Luft- und Raumfahrttechnik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
25244	Berufe in der Schiffbau-technik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
25254	Berufe in der Zweirad-technik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
34234	Berufe in der Kältetechnik - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
34304	Berufe in der Ver- und Entsorgung (ohne Spezialisierung) - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure
34344	Berufe im Anlagen-, Behälter- und Apparatebau - hoch komplexe Tätigkeiten	2144	Maschinenbauingenieure

Quelle: BA, 2011b

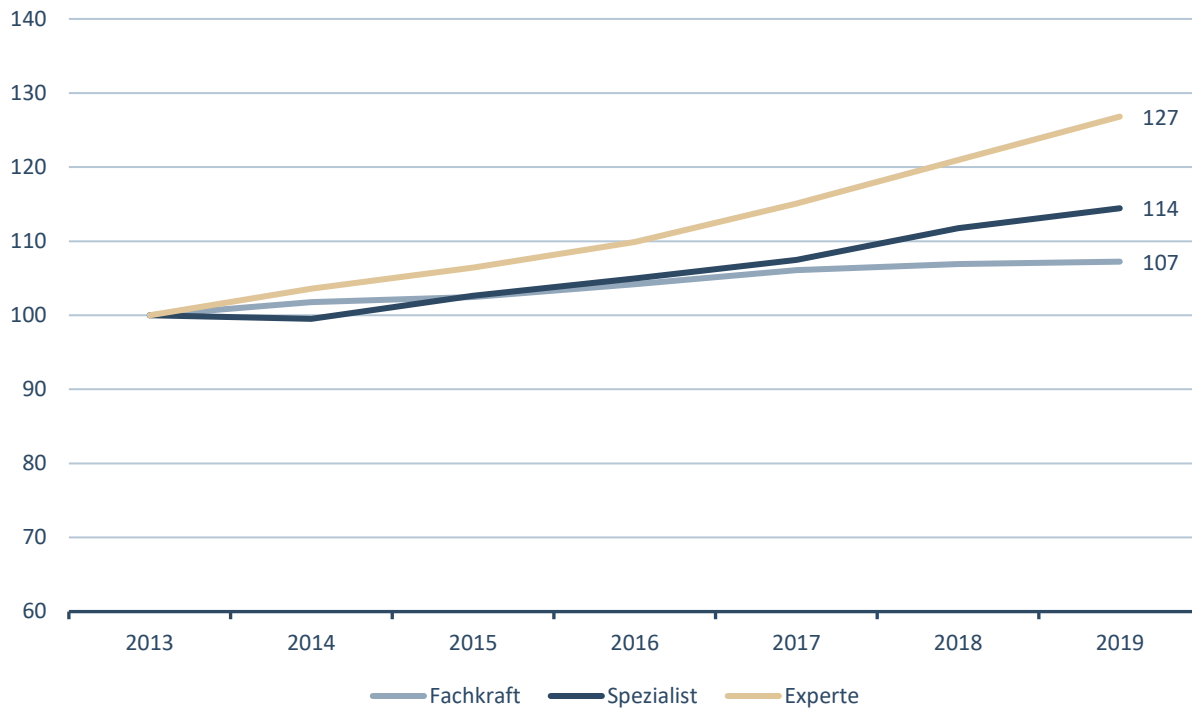
Die ISCO-Klassifikation bildet die Berufestruktur in Deutschland nicht differenziert genug ab und erscheint weniger geeignet als die KldB 2010, um für die Automobil- und Zulieferindustrie relevante Berufe zu identifizieren. Daher wird das in Kapitel 2.1.1 beschriebene Verfahren – die gemeinsame Betrachtung von Berufen (KldB 2010) und Branchen (WZ 2008) – favorisiert, um die relevanten Berufe zu ermitteln.

2.3 Gruppierung der Berufe anhand steigender und sinkender Nachfrage

Nach Abgrenzung der relevanten Berufsgattungen werden weitergehende Analysen durchgeführt, um den Stellenwert der jeweiligen Berufsgattung im Strukturwandel zu bestimmen. Hierzu gehören zum Beispiel Zeitreihenbetrachtungen der sozialversicherungspflichtig Beschäftigten, um Berufsgattungen mit steigender und sinkender Nachfrage zu identifizieren. Über die Arbeitslosen- und Stellenstatistik der BA können zudem weitere Informationen zur Fachkräftesituation in den jeweiligen Berufen dargestellt werden wie die Engpassrelation oder auch die aktuelle Fachkräftelücke (für weitere Informationen zu Methodik und Kennzahlen siehe Burstedde et al., 2020). Verschiedene Indikatoren geben dabei Aufschluss über aktuelle Entwicklungen des Arbeitsmarktes.

Dies soll beispielhaft für die Fachkräfte, Spezialisten und Experten der Fahrzeugtechnik dargestellt werden. Abbildung 2-2 zeigt die Beschäftigtenentwicklung in diesen drei Berufen für Baden-Württemberg. Auf allen drei Niveaus ist ein Beschäftigungsaufbau seit 2013 zu erkennen. Am stärksten ausgeprägt ist der Anstieg auf dem Niveau der Experten.

Abbildung 2-2: Veränderung der Beschäftigung in den drei Kraftfahrzeugtechnik-Berufen in Baden-Württemberg (Normiert 2013 = 100)



Quelle: IW-Berechnungen auf Basis von Sonderauswertungen der Bundesagentur für Arbeit, 2020

Eine Auswahl verschiedener Engpassindikatoren für diese Berufe zeigt Tabelle 2-4. Die unterschiedlichen Analysen dienen dazu, die identifizierten Berufe in Gruppen einzuteilen: Berufe mit steigender und Berufe mit sinkender Nachfrage. Diese Berufsgruppen können in der Ähnlichkeitsanalyse gegenübergestellt werden (siehe Kapitel 4).

Tabelle 2-4: Ausgewählte Engpass-Indikatoren in den drei Kraftfahrzeugberufen für Baden-Württemberg

Stand: 30.06.2020

KldB 5-Steller	Berufsuntergruppe	Anforderungsniveau	Engpassrelation	Fachkräftelücke	Stellenüberhangsquote	Offene Stellen
25212	Kraftfahrzeugtechnik	Fachkraft	36*	1.394	65	2.160
25213	Kraftfahrzeugtechnik	Spezialist	126	0	0	14
25214	Kraftfahrzeugtechnik	Experte	45*	103	55	189

Quelle: IW-Berechnungen auf Basis von Sonderauswertungen der Bundesagentur für Arbeit und des IAB, 2020

Hinweise: Die Engpassrelation ist das Verhältnis von Arbeitslosen zu offenen Stellen. Kommen weniger als 100 passend qualifizierte Arbeitslose auf 100 offene Stellen, liegt ein Engpass (*) vor. Die Fachkräftelücke ist die Zahl der offenen Stellen abzüglich der Arbeitslosen in einem Beruf. Die Stellenüberhangsquote ist der Anteil der Fachkräftelücke an allen offenen Stellen in einem Beruf. Die offenen Stellen ergeben sich aus den bei der BA gemeldeten Stellen durch die Hochrechnung mit Meldequoten und unter anteiliger Berücksichtigung von Zeitarbeitsstellen. Weitere Informationen zur IW-Methodik in Burstedde et al. (2020).

3 Analyse der beruflichen Kompetenzen in der Automobil- und Zulieferindustrie

Auf Grundlage der in Schritt 1 (Kapitel 2) identifizierten Berufe soll in Schritt 2 die Analyse der Kompetenzen in den jeweiligen Berufen erfolgen. Hierfür können verschiedene Quellen ausgewertet werden, die Informationen zu den Kompetenzen in den jeweiligen Berufen enthalten und mit der KldB 2010 kompatibel sind. Zu den Quellen gehören das Berufenet der BA, Aus- und Fortbildungsordnungen oder auch die Europäische Klassifikation für Fähigkeiten, Kompetenzen, Qualifikationen und Berufe (ESCO). Zudem kann die anschließende Analyse der Stellenanzeigen (siehe Kapitel 3.2) tiefere Einblicke in aktuelle Kompetenzanforderungen einzelner Berufsgattungen seitens der Unternehmen geben.

3.1 Analyse von Berufsinformationen

Im Folgenden werden die verschiedenen Quellen zur Beschreibung von Kompetenzen in verschiedenen Berufen kurz dargestellt und ihre Vor- und Nachteile aufgezeigt. Da Berufe grundsätzlich in unterschiedlichen Branchen ausgeübt werden können, enthalten die Berufsbeschreibungen häufig branchenübergreifende Tätigkeiten und Kompetenzen. Daher ist es denkbar, dass ein Teil der spezifischen Tätigkeiten, die in der Automobil- und Zulieferindustrie von besonderer Bedeutung sind, in den Berufsbeschreibungen nicht explizit genannt wird. Auch um diese Lücke schließen zu können und ebenfalls spezifische Kompetenzen erfassen zu können, wird eine Analyse von Stellenanzeigen empfohlen (siehe Kapitel 3.2).

3.1.1 Berufenet, Aus- und Fortbildungsordnungen sowie Studienordnungen

Ausführliche Informationen über Berufe in Deutschland finden sich in Berufenet der Bundesagentur für Arbeit (BA). Berufenet enthält 3.770 Berufe, die sich durch eine Systematiknummer der entsprechenden KldB 2010 zuordnen lassen. In der Datenbank werden für Berufe aller Qualifikationsniveaus Tätigkeiten aufgelistet. Die Tätigkeitsbeschreibung der einzelnen Berufe lässt Rückschlüsse auf die benötigten Kompetenzen zu. Die Informationen zu den einzelnen Berufen in Berufenet werden monatlich überprüft und bei Bedarf aktualisiert.

Für Ausbildungsberufe bietet Berufenet zusätzlich eine kompakte Übersicht über die Ausbildungsinhalte, in der beschrieben wird, welche Kompetenzen für den Beruf benötigt und erlernt werden. Da in der Regel jeder Ausbildungsberuf einer Berufsgattung und damit dem später ausgeübten Beruf zugeordnet werden kann, ist die Übereinstimmung zwischen den vermittelten Ausbildungsinhalten und den später benötigten Kompetenzen sehr hoch. Die Beschreibung der Ausbildungsinhalte basiert auf den Ausbildungsordnungen und wird in Berufenet ergänzt durch die konkrete Nennung von Arbeitsmaterialien und -techniken. Ein Beispiel dafür ist die Nutzung von CAD-Systemen bei dem Beruf des Karosserie- und Fahrzeugbaumechanikers, die in Berufenet konkret genannt wird, während in der Ausbildungsordnung nur „Konstruieren“ als Kompetenz genannt wird. Änderungen in den Ausbildungsordnungen werden zeitnah in Berufenet übernommen. Auch für Berufe, die einen Fortbildungsabschluss voraussetzen, wie Meister, Techniker oder Fachwirt, werden die Inhalte der Fortbildungsordnung in Berufenet integriert und regelmäßig aktualisiert.

Um berufliche Kompetenzen darzustellen, könnten alternativ oder ergänzend zum Berufenet für Fachkräfte und Spezialisten auch Aus- und Fortbildungsordnungen ausgewertet werden. Dabei ist zu beachten, dass Ausbildungsordnungen gestaltungsoffen und technologieneutral formuliert sind. Die technologieneutrale Formulierung der Ausbildungsordnungen hat den Vorteil, dass Ausbildungsordnungen mit technologischen Entwicklungen Schritt halten können, ohne jährlich aktualisiert werden zu müssen. Der Nachteil für das vorliegende Projekt ist jedoch, dass häufig genutzte neue Technologien, wie zum Beispiel Industrie-4.0-Technologien, keine explizite Erwähnung finden. Da Berufenet die Informationen aus den jeweiligen Aus- und Fortbildungsordnungen zusammen mit den übrigen Informationen zu Materialien und Techniken zu einem Beruf bereitstellt, würde eine Analyse beider Quellen höchstwahrscheinlich zu einer Doppelung der gefundenen Kompetenzen führen. Zudem wären die Kompetenzprofile auf Basis von Aus- und Fortbildungsordnungen weniger konkret.

Für Berufe, die ein Studium voraussetzen, ist die Zuordnung von Studienfach zu einer Berufsgattung nicht eindeutig. Für die Tätigkeit als Ingenieur für Fahrzeugtechnik beispielsweise kann man sich durch das Absolvieren unterschiedlicher Studiengänge qualifizieren (Automobilproduktion und Technik, Fahrzeug- und Antriebstechnik oder Automotive Systems Engineering). Um Kompetenzprofile für eine Tätigkeit in einer Berufsgattung zu erstellen, ist es daher nicht sinnvoll, alle Studienordnungen der möglichen Studienfächer zu überprüfen, weil dort viele Kompetenzen enthalten sind, die für die konkrete Tätigkeit in der betrachteten Berufsgattung nicht relevant sind. Hinzu kommt, dass es eine große Bandbreite an Studienordnungen gibt, die zum Teil von Hochschule zu Hochschule variieren. Daher wird von einer ergänzenden Auswertung der Studienordnungen abgeraten.

Für das Projekt empfehlen wir, die Kompetenzprofile aus Berufenet abzuleiten, da dort für alle Berufe Informationen zu Tätigkeiten und Kompetenzen vorliegen. Es ist zu prüfen, inwieweit die Auswertung von Berufenet händisch oder (halb-)automatisiert, beispielsweise über die Erstellung von Wörterbüchern, erfolgen kann, um die Kompetenzen bestmöglich abzubilden und gleichzeitig eine effiziente Arbeitsweise zu ermöglichen.

3.1.2 ESCO

Geprüft wurde ebenfalls die Nutzung von ESCO, die Europäische Klassifikation für Fähigkeiten, Kompetenzen, Qualifikationen und Berufe (Europäische Kommission, 2020). ESCO beschreibt, identifiziert und klassifiziert Berufe anhand von Tätigkeiten und Kompetenzen, die für den EU-Arbeitsmarkt und den Bereich der allgemeinen und beruflichen Bildung relevant sind. Der Datensatz steht in einem Online-Portal zur Verfügung, und die Informationen können kostenlos eingesehen und heruntergeladen werden. In ESCO werden zu jedem Beruf Informationen über den Inhalt der Tätigkeit sowie benötigte Kompetenzen zur Verfügung gestellt. Auch wenn die ESCO-Datenbank auf der ISCO-08-Klassifikation beruht, enthält sie Informationen zu 2.942 Einzelberufen. Hierfür wird die ISCO-Klassifikation noch in weitere spezifische Berufe gegliedert.

Über den Umsteigeschlüssel der BA zwischen der KldB 2010 und ISCO-08 (BA, 2011b) kann auf die ISCO-08-4-Steller geschlossen werden und innerhalb der 4-Steller könnte dann geprüft werden, ob die passenden, spezifischen Berufe hinterlegt sind. Damit enthält die ESCO-Datenbank eine fünfte Ebene mit spezifischen Berufen, die vergleichbar mit der fünften Ebene der KldB

2010 ist. Es gibt jedoch zwei Herausforderungen bei der Verwendung der ESCO-Datenbank: zum einen die nicht immer eindeutige Zuordnung von KldB-5-Stellern zu ISCO-08-Berufen und zum anderen die nicht immer vorhandene Differenzierung der spezifischen Berufe in der ESCO-Datenbank nach Qualifikationsniveau. So gibt es Berufe, bei denen der Umstieg nicht eindeutig ist. In diesem Fall werden einem KldB-5-Steller mehrere ISCO-08-Codes zugeteilt. Dies ist auch für einige der relevanten Berufe der Automobilindustrie der Fall, zum Beispiel bei den Maschinen- und Anlageführer/innen sowie bei den technischen Servicekräften in Wartung und Instandhaltung. Zudem kommt es vor, dass 5-Steller der KldB 2010 der Qualifikationsniveaus 2 (Fachkräfte) und 3 (Spezialisten) demselben ISCO-08-4-Steller zugeordnet sind. Die Differenzierung von Kompetenzen unterschiedlicher Qualifikationsniveaus ist jedoch für den deutschen Arbeitsmarkt und die Qualifizierungswege bedeutsam. Für die Vergleichbarkeit von Berufen sollte klar sein, ob es sich um einen Beruf handelt, der eine (duale) Ausbildung voraussetzt (Niveau 2) oder einen Fortbildungsabschluss wie einen Meister, Techniker oder Fachwirt oder einen Bachelorabschluss (Niveau 3).

Auch wenn die ESCO-Datenbank detaillierte Informationen über Tätigkeiten und Kompetenzen liefert, befürworten wir die Verwendung von Berufenet, weil eine eindeutige Zuordnung der Berufe zur KldB 2010 aus unserer Sicht unerlässlich ist. Berufenet enthält alle auf dem deutschen Arbeitsmarkt relevanten Berufsabschlüsse und spiegelt das deutsche Berufsbildungssystem mit seinen Besonderheiten wider (so haben beispielsweise Fortbildungsabschlüsse in vielen Ländern nicht den gleichen Stellenwert wie in Deutschland). Des Weiteren sollte die Gruppierung der relevanten Berufe anhand steigender und sinkender Nachfrage auf Basis der KldB 2010 erfolgen, denn nur auf Basis der KldB 2010 liegen detaillierte Daten über die Beschäftigungsentwicklung vor. Da die Kompetenzprofile, die aus Berufenet für die relevanten Berufe abgeleitet werden, durch die Analyse der Stellenanzeigen erweitert werden (Kapitel 3.2), kann die Gefahr, dass wichtige Kompetenzen unberücksichtigt bleiben, minimiert werden.

3.1.3 Weitere Datenquellen

Es wurde auch geprüft, ob weitere Datenquellen wie beispielsweise die BIBB-BAuA-Erwerbstätigenbefragung 2018 zur Identifizierung von Kompetenzen und Tätigkeitsschwerpunkten in der Automobil- und Zuliefererindustrie herangezogen werden können. Trotz einer Stichprobengröße von 20.012 Erwerbstätigen lassen sich Branchen und Berufe für das vorliegende Forschungsinteresse nicht fein genug abgrenzen. Im Scientific Use File werden die Berufe nach KldB 2010 nur auf 2- beziehungsweise 3-Steller-Ebene zur Verfügung gestellt (Abbildung 2-1). Im Rahmen eines Gastwissenschaftleraufenthalts kann Zugang zu den anonymisierten Daten auf KldB 5-Steller-Ebene beantragt werden. Allerdings sind die Fallzahlen bereits bei Aufteilung nach groben Wirtschaftsbereichen relativ gering, der Bereich Industrie umfasst zum Beispiel nur noch 3.880 Beobachtungen. Damit sind die Fallzahlen bei einer feineren Abgrenzung nicht mehr ausreichend für eine aussagekräftige Analyse (Rohrbach-Schmidt/Hall, 2020).

3.2 Analyse von Stellenausschreibungen der Automobil- und Zulieferindustrie

Komplementär zu der in Kapitel 3.1 diskutierten Analyse von Kompetenzen der Berufe in der Automobil- und Zulieferindustrie können (Online-)Stellenanzeigen einen weiteren wichtigen Baustein darstellen, um ein umfassendes Kompetenzprofil der relevanten Berufe zu erstellen.

Die Vorteile von Stellenanzeigen bei der Identifikation relevanter berufsspezifischer Kompetenzen gegenüber traditionellen Informationsquellen belaufen sich vor allem auf ihre Aktualität, ihren Umfang und ihre Spezifität (siehe Stops et al., 2020; Turrel et al., 2019). So können Online-Stellenanzeigen tagesaktuelle Informationen zu von Arbeitgebern gesuchten Kompetenzen liefern. Auch die konkrete Ausformulierung der gewünschten Kompetenzen kann einen Vorteil gegenüber gebräuchlichen Klassifikationen von Kompetenzen bieten, die die Anforderungen notwendigerweise auf höheren, technologieunabhängigen Klassifikationsebenen zusammenfassen. Darüber hinaus beinhalten Stellenanzeigen fast immer weitere wichtige Informationen, wie etwa die Adresse, die Region oder den Namen des ausschreibenden Unternehmens, die für weiterführende sozio-ökonomische Fragestellungen mit erhoben und genutzt werden können.

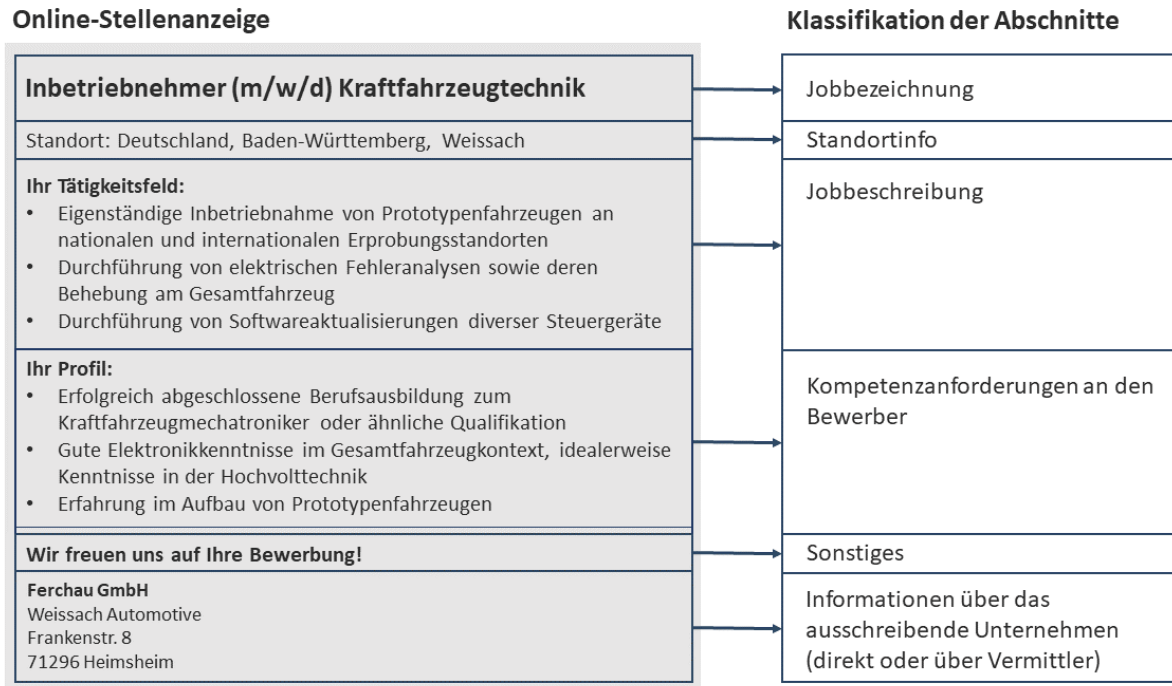
Dennoch können Online-Stellenanzeigen kein umfassendes Bild der Nachfrage auf dem deutschen Arbeitsmarkt darstellen. So arbeiten Hershbein und Kahn (2018) heraus, dass Online-Stellenanzeigen Stellengesuche für gut ausgebildete Arbeitnehmer tendenziell überrepräsentieren. Auch Cajner et al. (2016) zeigen Diskrepanzen zwischen der Anzahl von online veröffentlichten Stellenanzeigen und offiziellen Statistiken in den USA auf. Nimmt man allerdings an, dass der Anteil an Stellenanzeigen, die online gestellt werden, stetig ansteigt (siehe Hershbein/Kahn 2018; Turrell et al., 2019), können Online-Stellenanzeigen vor allem in der komplementären Nutzung mit anderen relevanten Quellen wie Berufsinformationen aus Berufenet eine wichtige zusätzliche Möglichkeit darstellen, ein möglichst umfassendes Bild an berufsspezifischen Kompetenzanforderungen zu erarbeiten. Hinzu kommt, dass die großen kommerziellen Anbieter von Daten zu Stellenanzeigen wie BurningGlass oder Textkernel nicht nur die Online-Stellenanzeigen der großen Jobportale sammeln, sondern auf weitere, mehrere zehntausende Quellen zurückgreifen, wie etwa Unternehmensseiten oder die Stellenbörsen traditioneller (Tages-)Zeitungen. Für Deutschland hat beispielsweise der Datenanbieter Textkernel für das Jahr 2019 über 12,5 Millionen eindeutige, um Duplikate bereinigte Stellenanzeigen erhoben, die durch das Scrapen von knapp 1.000 Jobportalen und über 60.000 Firmenkarriereseiten sowie Zeitungen generiert wurden.⁴

Abbildung 3-1 zeigt beispielhaft den Inhalt einer Online-Stellenanzeige für den Beruf des „Inbetriebnehmers (m/w/d) Kraftfahrzeugtechnik“. Die Klassifikation der Abschnitte der Stellenanzeige macht deutlich, welche Informationen grundsätzlich aus dem Volltext einer Anzeige generiert werden können. Die für die ausgeschriebenen Stellen geforderten Kompetenzen sind im Abschnitt „Kompetenzanforderungen an den Bewerber“ zu finden, aber auch im Abschnitt „Jobbeschreibung“ ist eine Nennung geforderter Kompetenzen möglich. Im Beispiel werden neben der für den Beruf relevanten Ausbildung, welche im Abschnitt der Kompetenzanforderungen genannt wird, in dem Abschnitt der Jobbeschreibung zusätzlich Kenntnisse der Inbetriebnahme von Prototypenfahrzeugen und der elektrischen Fehleranalyse gefordert. Weitere wichtige Informationen, die für detailliertere Analysen genutzt werden können, sind die Bezeichnung des Jobs und der Standort beziehungsweise die Adresse des Arbeitgebers (siehe hierzu Kapitel 3.2.1.1 und 3.2.1.2). Da die angegebene Adresse auch der eines Job-Vermittlers entsprechen

⁴ Zu beachten ist hier das Volumen der jeweiligen Datensätze. So haben die Daten der Stellenausschreibungen (mit Duplikaten) für 2020 von Textkernel ein Gesamtvolumen von ca. 360 Gigabyte. Hier müssen die entsprechende Hardware beziehungsweise notwendige Big Data Tools vorhanden sein, um die Daten verarbeiten und vor allem modellieren zu können.

kann, muss hier teilweise unterschieden werden zwischen dem eigentlichen Beschäftigungsort und der angegebenen Unternehmensadresse. Auf diese Weise ist es möglich, Stellenanzeigen für die baden-württembergische Automobilindustrie zu identifizieren.

Abbildung 3-1: Beispielhafte Online-Stellenanzeige für einen „Inbetriebnehmer (m/w/d) Kraftfahrzeugtechnik“



Quelle: Indeed; eigene Darstellung

3.2.1 Identifikation von relevanten Stellenausschreibungen der Automobil- und Zulieferindustrie

Damit die Informationen, die aus Stellenanzeigen extrahiert werden, maximal passgenau für die Fragestellung möglicher Übergangspfade in der Automobil- und Zulieferindustrie in Baden-Württemberg sind, müssen zunächst die relevanten Ausschreibungen im gesamten Korpus der Stellenanzeigen identifiziert werden. Anschließend können die darin enthaltenen Kompetenzen identifiziert, deskriptiv aufgearbeitet und für weitere Analysen verwendet werden. Im Folgenden werden drei Ansätze diskutiert, mit deren Hilfe die relevanten Stellenanzeigen identifiziert werden können. Diese können unabhängig voneinander eingesetzt werden. Für ein möglichst passgenaues Ergebnis bietet sich jedoch eine Kombination der Ansätze an.

3.2.1.1 Identifikation relevanter Stellenausschreibungen über die Bezeichnung der Berufe

Eine erste Möglichkeit, die relevanten Stellenausschreibungen zu identifizieren, basiert auf den Informationen, die in der Stellenbezeichnung beziehungsweise dem genannten Beruf der ausgeschriebenen Stelle angegeben sind. Diese Berufsbezeichnungen können mit Hilfe eines anzulegenden Wörterbuchs durchsucht werden, um die relevanten Stellenausschreibungen zu

identifizieren.⁵ Als Inhalte des Wörterbuchs können die in Kapitel 2 erarbeiteten Berufsbezeichnungen aus der Automobil- und Zulieferindustrie genutzt werden. Für das Wörterbuch werden diese Berufsbezeichnungen um reguläre Ausdrücke, wie alternative Schreibweisen, ergänzt, um möglichst viele relevante Ausschreibungen identifizieren zu können.⁶ Hierfür kann beispielsweise auch das Schlüsselverzeichnis der BA verwendet werden, welches Arbeitgeber im Rahmen des Meldeverfahrens zur Sozialversicherung nutzen (siehe Kapitel 2.1).

Um weitere relevante Ausschreibungen identifizieren zu können, bietet es sich an, ein iteratives Verfahren anzuwenden. Hierbei wird das bestehende Wörterbuch als initiale Suchliste verwendet. Mit Hilfe der in diesem Wörterbuch hinterlegten Begriffe werden über die Berufsbezeichnung aller Stellenausschreibungen automatisiert relevante Stellenausschreibungen identifiziert. Die Volltexte dieser Stellenanzeigen werden händisch nach relevanten Schlagwörtern durchsucht. Diese Schlagwörter werden in ein weiteres separates Wörterbuch abgespeichert. Unter Verwendung dieses zweiten Wörterbuchs werden die Volltexte aller Stellen erneut durchsucht. Berufsbeschreibungen der Stellenausschreibungen, die Begriffe dieses zweiten Wörterbuches enthalten, werden nach einer händischen Validierung in das erste Wörterbuch der Berufsbezeichnungen integriert. Anschließend können die Volltexte dieser neu identifizierten Stellenausschreibungen erneut nach relevanten Schlagwörtern durchsucht werden, und der iterative Prozess beginnt von vorne.⁷

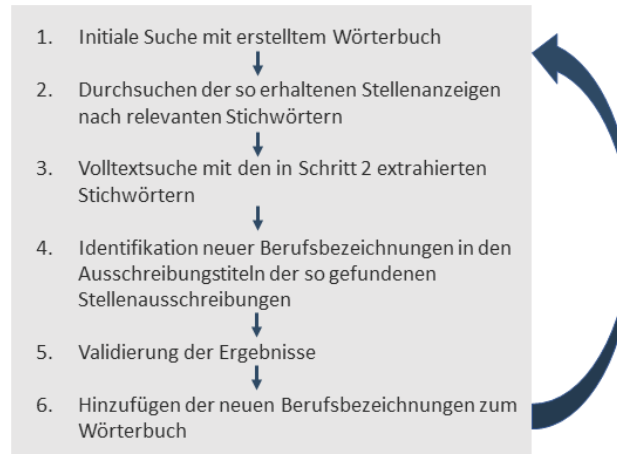
Durch eine wiederholte Durchführung dieser Arbeitsschritte können das Wörterbuch der Berufsbezeichnungen stetig erweitert und dadurch eine möglich große Anzahl relevanter Stellenanzeigen ermittelt werden. Dieser Prozess, dargestellt in Abbildung 3-2, wird solange wiederholt, bis keine weiteren relevanten Stellenanzeigen mehr gefunden werden, oder bis die Zeitkosten den Nutzen durch den Zugewinn neuer Stellenanzeigen überschreiten.

⁵ Siehe beispielhaft Demary et al. (2020, 21 f.) für die Identifikation von Stellenausschreibungen mit Bezug zu künstlicher Intelligenz aus einem umfassenderen Korpus an Stellenausschreibungen.

⁶ So kann etwa statt des Begriffs „elektrische Fehleranalyse“ der um reguläre Ausdrücke (regular expressions) ergänzte Begriff `elektrische[rnm]?[\-_,;\(\)\[\]\/\.\.]?fehleranalyse` genutzt werden. Damit werden mögliche linguistische Variationen des Begriffs wie „elektrischer Fehleranalysen“, „elektrische Fehleranalysemodule“ oder „(elektrische)-Fehleranalyse“ ebenfalls erkannt, die mit der einfachen Suche des Begriffs „elektrische Fehleranalyse“ nicht erkannt würden.

⁷ Da dieser Prozess zeitaufwendig ist, können Text Mining Verfahren genutzt werden, die den Prozess beschleunigen können. Siehe Kapitel 3.2.1.3 für eine Beschreibung der Verfahren und Kapitel 3.2.2.1 für die Diskussion der Kombination der Verfahren mit Wörterbüchern.

Abbildung 3-2: Iterativer Prozess der Identifikation von Stellenanzeigen über eine Wörterbuchsuche mit Berufsbezeichnungen



Quelle: Eigene Darstellung

3.2.1.2 Identifikation relevanter Stellenausschreibungen über Unternehmensname

Alternativ oder auch zusätzlich kann eine Liste mit den in Baden-Württemberg ansässigen Unternehmen der Automobil- und Zulieferindustrie erstellt werden. Hierbei muss unter Einbindung der Stakeholder entschieden werden, wie weit die Automobil- und Zulieferindustrie gefasst wird, ob die Abgrenzung beispielsweise nach der in Kapitel 2.1 als relevant identifizierten Abteilungen der Klassifikation der Wirtschaftszweige erfolgt oder darüber hinaus auch Unternehmen im Sinne eines zugehörigen „Automobilclusters“ betrachtet werden sollen (zur Identifikation von Automobilclustern siehe die Arbeiten der IW Consult im Projekt „Wirtschaftliche Bedeutung regionaler Automobilnetzwerke in Deutschland“ im Auftrag des BMWi). Gegeben dieser Entscheidung können die entsprechenden Unternehmen über Unternehmensdatenbanken identifiziert werden (beispielsweise über die Firmendatenbank des Baden-Württembergischen Industrie- und Handelskammertages oder den Datenanbieter beDirect). Hierbei muss keine Vollerhebung der Unternehmen der Automobil- und Zuliefererindustrie in Baden-Württemberg erfolgen, da Ausbildungsverordnungen gleich und Stellenausschreibungen für gleiche Tätigkeiten ab einer ausreichend großen Anzahl an Stellenausschreibungen ähnlich sein sollten. In Kooperation mit den relevanten Stakeholdern sollte die entstandene Unternehmensliste ergänzt und validiert werden.

Mit Hilfe dieser Unternehmensnamen können über den Abschnitt der Stellenausschreibung, der die Information des ausschreibenden Unternehmens beinhaltet, oder über die Volltexte (weitere) relevante Stellenausschreibungen identifiziert werden.

Wird die Wörterbuchsuche mit Berufsbezeichnungen, die in 3.2.1.1 beschrieben wurde, vorrangig durchgeführt, können zudem die Unternehmensnamen aus den dort identifizierten Stellenausschreibungen in die Unternehmensliste mit eingehen. Hierdurch können zu den in 3.2.1.1 identifizierten Stellenausschreibungen weitere relevante gefunden werden, da über diesen Ansatz alle Stellenausschreibungen des jeweiligen Unternehmens in den Korpus eingehen.

Die Stellenausschreibungen, die so identifiziert werden, sind zunächst nur potenziell relevante Stellenausschreibungen. Dies liegt daran, dass Unternehmen der Automobil- und Zuliefererindustrie zum Beispiel mit der Suche nach Reinigungskräften ebenso Stellen ausschreiben, die nicht zu den Berufen gehören, die im Fokus der Übergangspfadanalyse sind (siehe auch Kapitel 2.1). Daher muss der Datensatz noch entsprechend bereinigt werden (siehe hierzu Abschnitt „Diskussion der Verfahren auf den Anwendungsfall“).

3.2.1.3 Identifikation relevanter Stellenausschreibungen mittels Machine Learning Verfahren

Neben den in Kapitel 3.2.1.1 und 3.2.1.2 diskutierten regelbasierten Ansätzen zur Klassifikation von Stellenanzeigen können auch Machine Learning Verfahren genutzt werden, um die relevanten Stellenausschreibungen der Automobil- und Zulieferindustrie zu identifizieren. Die Nachteile regelbasierter Ansätze sind einerseits der hohe Arbeits- und Zeitaufwand für die manuelle Erstellung der Regeln (siehe Hermes/Schandock, 2016), andererseits liefern state-of-the-art Machine und Deep Learning Modelle in vielen Anwendungsbereichen seit einigen Jahren zuverlässig bessere Klassifikationsergebnisse als regelbasierte Verfahren (Amato et al., 2015; Boselli et al., 2018; Colombo, 2018; Minaee et al., 2020).

Machine Learning Verfahren lassen sich grundsätzlich in überwachte (supervised) und unüberwachte (unsupervised) Verfahren unterteilen. Für die Klassifikation von Textdaten werden meist überwachte Verfahren eingesetzt, doch auch unüberwachte Verfahren können hierfür angewendet werden (Amato et al. 2015; Minaee et al. 2020). Für überwachte Ansätze muss zunächst ein Trainingsdatensatz mit händisch klassifizierten Ausschreibungsvolltexten verschiedener Branchen erstellt werden. Den Stellenanzeigen im Trainingsdatensatz wird ein Label zugewiesen, welches angibt, ob es sich um eine Ausschreibung der Automobil- und Zulieferindustrie handelt oder nicht. Diese Trainingsdaten werden dann in ein Machine Learning Modell gegeben, in welchem ein Klassifikator (classifier) versucht, auf Grundlage dieser manuell klassifizierten Daten vorherzusagen, ob eine bisher unbekannte Stellenanzeige der Automobil- und Zulieferindustrie zugeordnet werden kann oder nicht. Im Folgenden soll eine Auswahl verschiedener Machine und Deep Learning Methoden kurz beschrieben und deren möglicher Nutzen für den konkreten Anwendungsfall evaluiert werden.⁸

Überwachte Machine Learning Verfahren mit Bag-of-Words Ansatz

Da fast alle Machine Learning Verfahren numerische Daten als Input benötigen, besteht ein erster Schritt zunächst darin, die Textdaten in ein entsprechendes Format zu transformieren. Für diesen Prozess wird häufig die sogenannte „Bag-of-Words“ Annahme getroffen. Dabei werden Textdaten als eine unsortierte Gruppe von Wörtern repräsentiert und lediglich die Häufigkeit eines jeden Wortes erfasst, unabhängig davon, an welcher Position das Wort in einem Satz vorkommt. Neben der Häufigkeit eines Wortes in einem Textdokument können weitere (teilweise vorgelagerte) Textbearbeitungsschritte durchgeführt und Metriken erhoben werden. So können etwa zunächst sogenannte Stopp-Wörter, dies sind zum Beispiel Artikel, die in der Regel sehr

⁸ Da eine ausführliche Diskussion der hier aufgezählten Machine und Deep Learning Verfahren im Rahmen dieser Expertise den vorgegebenen Umfang deutlich überschreiten würde, werden die jeweiligen Verfahren lediglich in ihren Grundzügen besprochen. Die angegebene Literatur bietet einen vertiefenden Einblick in die jeweiligen Ansätze.

häufig vorkommen aber keine inhaltliche Bedeutung haben, entfernt werden. Zusätzlich können Wörter auf ihren Wortstamm reduziert werden (Lemmatisierung) oder anderweitig normalisiert werden (etwa durch durchgehende Kleinschreibung oder das Zusammenfassen beziehungsweise Entfernen von Zahlen und Sonderzeichen). Zuletzt können auch Bi- oder n-Gramme gebildet und gezählt werden. Hierbei werden jeweils zwei oder mehr Wörter zusammen erfasst und gezählt, um mögliche häufig vorkommende Wortkombinationen abbilden zu können (siehe Hermes/Schandock, 2016; Jurasky/Martin, 2019).

Zählt man in einem weiteren Schritt alle Wörter oder Wortkombinationen eines jeden Dokuments, lässt sich die erhaltene Worthäufigkeit als Vektor darstellen. Dieser Vektor gibt an, wie häufig ein Wort aus der Gesamtheit an Wörtern aus allen Stellenanzeigen in dem jeweiligen Dokument (hier der jeweiligen Stellenanzeige) vorkommt. Die Gesamtheit aller Vektoren kann in einer Häufigkeitsmatrix dargestellt werden (siehe Abbildung 3-3).

Abbildung 3-3: Beispielhafte Häufigkeitsmatrix für alle Wörter im Gesamtkorpus aller Stellenanzeigen

Wörter	Stellenanzeige 1	Stellenanzeige 2	Stellenanzeige 3	...	Stellenanzeige n
anforderung	2	4	3	...	6
auslegung	1	0	0	...	0
...
zellmodul	5	4	0	...	0

Quelle: Eigene Darstellung

Ein Nachteil der dargestellten Häufigkeitsmatrix ist allerdings der Umstand, dass die dargestellten Worthäufigkeiten in vielen Fällen nicht ausreichen, um effizient zwischen Texten verschiedenen Inhalts zu unterscheiden. Nimmt man das Beispiel der Stellenanzeigen, ist es sehr wahrscheinlich, dass viele Ausschreibungen die Wörter „Anforderungen“, „Kenntnisse“ oder „Erfahrung“ beinhalten. Um aber zwischen verschiedenen Berufen zu unterscheiden, sind spezifischere Begriffe wie etwa „HV-Batterie“ oder „Zellmodul“ wesentlich entscheidender. Deshalb wird oft ein weiterer Bearbeitungsschritt durchgeführt, in welchem eine sogenannte „term frequency – inverse document frequency“ (tfidf) erstellt wird. In die Berechnung dieser Gewichtung eines Wortes geht neben der absoluten Häufigkeit eines Wortes pro Dokument auch die inverse Dokumenthäufigkeit mit ein. Der tfidf Wert eines Wortes ist umso höher, je häufiger das Wort in wenigen Dokumenten vorkommt.⁹ Dadurch bekommen Wörter, die in weniger Dokumenten vorkommen, also ein höheres potenzielles Unterscheidungsgewicht zwischen Dokumenten haben, einen höheren Wert (Jurasky/Martin, 2019, 105 ff.). Das Beispiel in Abbildung 3-4 zeigt, dass so etwa das Wort „Zellmodul“, das lediglich in den Stellenanzeigen 1 und 2

⁹ Mathematisch lässt sich das Gewichtsmaß der term frequency – inverse document frequency (w) für jedes Wort (i) in Dokument (j) darstellen als $w_{ij} = tf_{ij} \times \log\left(\frac{N}{df_i}\right)$, wobei tf die Worthäufigkeit (term frequency) jedes Wortes in jedem Dokument angibt, N die Anzahl der Dokumente und df die Anzahl aller Dokumente, in denen Wort i genannt wird.

vorkommt, ein deutlich höheres Gewicht bekommt als das Wort „Anforderung“, welches in allen Stellenanzeigen vorkommt.

Abbildung 3-4: Beispielhafte tfidf-Matrix für alle Wörter im Gesamtkorpus der Stellenanzeigen

Wörter	Stellenanzeige 1	Stellenanzeige 2	Stellenanzeige 3	...	Stellenanzeige n
anforderung	0.74	0.95	0.87	...	1.03
auslegung	0.65	0	0	...	0
...
zellmodul	2.87	2.45	0	...	0

Quelle: Eigene Darstellung

Mit der numerischen Repräsentation der Textdaten lassen sich in einem weiteren Schritt verschiedene Machine Learning Klassifikatoren anwenden, um von den manuell codierten Trainingsdaten auf den um ein Vielfaches größeren Gesamtdatensatz zu schließen und die relevanten Stellenausschreibungen zu identifizieren. Häufig angewendete Machine Learning Klassifikatoren sind etwa:

- *Naive Bayes*: Der Naive-Bayes Klassifikator ist ein bayesianischer Ansatz, der sehr vereinfachende (naive) Annahmen darüber trifft, wie Merkmale (Features), im Beispiel von Stellenanzeigen die Wörter, in einem Datensatz zusammenhängen. Für Textdaten wird, wie oben beschrieben, die Bag-of-Words Annahme getroffen. Über die Häufigkeit der in den Dokumenten vorkommenden Wörter gibt der Klassifikator unter Verwendung des Satz von Bayes eine Wahrscheinlichkeit an, mit der ein Dokument zu einer bestimmten Klasse gehört (siehe Jurasky/Martin, 2019; McCallum/Nigam, 1998; Rennie et al., 2003).
- *Support Vector Machines*: Ein weiterer Ansatz, der häufig für die Klassifikation von Textdaten genutzt wird, sind sogenannte Support Vector Machines (SVM). SVM sind lineare Klassifikatoren, die versuchen, zwischen Klassen zu unterscheiden, beispielsweise ob eine Stellenanzeige der Automobil- und Zulieferindustrie zuzuordnen ist oder nicht. Hierzu wird eine Hyperebene identifiziert, die die Distanzen der oben beschriebenen Vektoren maximiert (Tong/Koller, 2001).
- *Random Forest*: Ein Alternative zu den oben genannten Ansätzen für die Klassifikation von Texten kann der sogenannte Random Forest Klassifikator darstellen. Dieser basiert auf mehreren unkorrelierten Entscheidungsbäumen, die während des Lernprozesses randomisiert werden (siehe Ali et al., 2012; Breiman, 2001). Dieser Klassifikator ist sehr effizient für größere Datenmengen und bietet gegenüber den oben genannten Ansätzen oft bessere Ergebnisse. Da sich die Güte der Ergebnisse aber von Fall zu Fall unterscheiden kann, werden üblicherweise mehrere Algorithmen beziehungsweise Klassifikatoren auf ein Klassifikationsproblem angewendet und deren Ergebnisse miteinander verglichen.

Unüberwachte Machine Learning Verfahren mit Bag-of-Words Ansatz

Neben überwachten Machine Learning Ansätzen bieten sich auch unüberwachte Verfahren an. Diese zeichnen sich dadurch aus, dass keine manuell codierten oder annotierten Trainingsdaten benötigt werden. Ein aus dieser Richtung vielfach angewendeter Ansatz sind sogenannte „Topic Models“ – eine Methode zur Verteilungsanalyse semantischer Wortgruppen in Textkorpora. Dieser Ansatz basiert auf der von Blei et al. (2003) vorgestellten „latent Dirichlet allocation“ (LDA).

LDA ist ein probabilistisches, bayesianisches Modell, welches auf der Grundannahme beruht, dass Texte auf den Häufigkeiten und den Wahrscheinlichkeiten bestimmter Wörter basieren. Konkret nimmt LDA an, dass jedes Dokument aus einem Mix aus verschiedenen Topics (Themen) besteht. Bezogen auf Stellenanzeigen könnte also eine Stellenanzeige zu 70 Prozent aus dem Topic „Automobilindustrie“, zu 20 Prozent aus dem Topic „Künstliche Intelligenz“ und zu 10 Prozent aus dem Topic „Sonstiges“ bestehen. Eine weitere Annahme ist, dass jedes Topic aus einem Mix an Wörtern zusammengesetzt ist. So könnten die häufigsten Wörter für das Topic „Automobilindustrie“ beispielhaft „Automobilbranche“, „Batterie“, und „Ingenieur“ sein, während das Topic „Künstliche Intelligenz“ „Machine Learning“, „Bilderkennung“ und „automatisiertes Fahren“ als häufigste Wörter hätte. Wichtig hierbei ist, dass dieselben Wörter in verschiedenen Topics vorkommen (können). Als Ergebnis können die Stellenanzeigen als relevant für die Automobil- und Zulieferindustrie deklariert werden, deren wahrscheinlichstes Topic das der Automobilindustrie ist.

Ein zentraler Parameter für das Trainieren von LDA Modellen ist die Anzahl an Topics (K), die im Vorhinein festzulegen ist. Weiß man etwa, dass der Textkorpus lediglich Stellenanzeigen aus drei verschiedenen Berufsgruppen enthält, würde der Wert für K entsprechend festgelegt. Weit- aus häufiger tritt allerdings der Fall ein, dass dieser Wert für K zunächst unbekannt ist. Hier sollten mehrere Modelle mit verschiedenen Werten für K trainiert und anhand verschiedener Diagnosekriterien evaluiert werden (siehe etwa Mimno et al., 2011).

Für LDA gibt es mittlerweile mehrere Alternativen für spezifischere Anwendungsfälle (beispielsweise für sehr kurze Texte). Ein nennenswerter Ansatz für die Klassifikation von Stellenanzeigen können die sogenannten „Structural Topic Models“ (Roberts et al., 2013) darstellen, da in diesem Ansatz neben den reinen Textdaten auch Kovariaten (beispielsweise das Datum der Stellenanzeige) auf Dokumentenebene (beispielsweise je Stellenausschreibung) mit in die Wahrscheinlichkeitsberechnung der Topics integriert werden können, um eine bessere Zuordnung des Topics zu erzielen.

Deep Learning Verfahren

Das Feld der natürlichen Sprachverarbeitung (NLP) ist sehr schnelllebig und hat in den letzten Jahren enorme Fortschritte gemacht. Vor allem der Bereich Deep Learning hat dabei immer mehr an Bedeutung gewonnen und erzielt mittlerweile in vielen Fällen sehr gute Ergebnisse in der Textklassifizierung. Dabei wird die Performance von traditionellen überwachten und unüberwachten Machine Learning Ansätzen (siehe oben) meist übertroffen (Minaee et al. 2020).

Ein großer Vorteil von Deep Learning Ansätzen ist die geringere Abhängigkeit von zeitintensivem Feature Engineering. Das bedeutet, dass viele der oben diskutierten Schritte zur Bereinigung, Standardisierung und Darstellung der Textdaten hier nicht vom Anwender übernommen werden müssen. Ein weiterer zentraler Vorteil von Deep Learning Modellen besteht in der adäquateren Repräsentation natürlicher Sprache. So wird hier nicht mit der Bag-of-Words Annahme gearbeitet, sondern Wörter können in ihrer grammatikalischen und semantischen Struktur und ihrem Kontext erfasst und verarbeitet werden. Dies bietet auch einen Vorteil gegenüber klassischen Word Embedding Ansätzen (wie etwa word2vec oder gloVe), die lediglich statische Vektorrepräsentationen von Wörtern enthalten. So kann beispielsweise im Deep Learning Verfahren „Bidirectional Encoder Representations from Transformers“ (BERT) von google das Wort „Bank“ in den Beispielsätzen „Ich hole Geld von der Bank“ und „Ich sitze auf einer Bank“ mit einer unterschiedlichen kontextualen Bedeutung verstanden werden (Devlin et al., 2019).

Des Weiteren bietet auch das sogenannte Pre-Training, welches im Zuge von Deep Learning Verfahren durchgeführt werden kann, einen Vorteil in der natürlichen Sprachverarbeitung gegenüber traditionellen Machine Learning Verfahren. In den Deep Learning Verfahren kann in einem unüberwachten Verfahren ein neuronales Netz auf einen großen Korpus nicht codierter Textdaten (vor-)trainiert werden. Hierdurch wird die Information über die kontextabhängige Sprache für den jeweiligen Anwendungsbereich verbessert. Da Deep Learning Modelle wie BERT mit unabhängigen Textdaten wie etwa dem Gesamtkorpus aller Wikipedia Artikel trainiert werden, kann es Sinn machen, das Modell mit einem spezifischen Textkorpus (zum Beispiel den der Stellenanzeigen) vorzutrainieren, damit die sprachlichen Eigenheiten dieser Text besser weiterverarbeitet werden können. Erst in einem darauffolgenden Schritt werden Trainingsdaten benutzt, um die Modelle zu optimieren (siehe Yang et al., 2019). Verschiedene Deep Learning Modelle wie BERT, ULMFit oder XLNet konnten in den letzten Jahren stetig bessere Ergebnisse bei verschiedenen NLP Problemen, darunter die Klassifikation von Texten, liefern (Minaee et al., 2020).

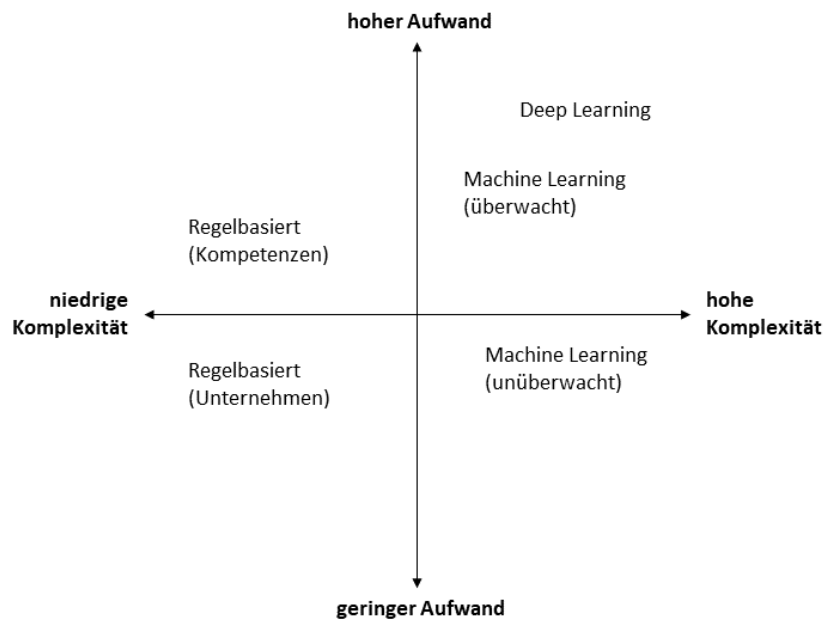
Diskussion der Verfahren auf den Anwendungsfall

Für die Identifikation relevanter Stellenausschreibungen der Automobil- und Zulieferindustrie in einem größeren Datensatz mit Stellenanzeigen können alle der oben diskutierten Ansätze einen Mehrwert darstellen. Da ex ante nicht abzusehen ist, welcher Ansatz die besten Ergebnisse liefern wird, bietet es sich an, verschiedene Ansätze auszuprobieren und diese eventuell zu synthetisieren.

In Abbildung 3-5 werden der angenommene Entwicklungsaufwand und die Komplexität der oben diskutierten Verfahren geschätzt. Der Entwicklungsaufwand umfasst dabei etwa die iterative Erstellung und Validierung der jeweiligen Wörterbücher bei den regelbasierten Ansätzen und die Erstellung der Trainingsdatensätze, die Bereinigung und Bearbeitung der Daten sowie die Entwicklung und Optimierung der jeweiligen Modelle bei den Machine und Deep Learning Ansätzen. Der geringere Arbeitsaufwand bei unüberwachten Machine Learning Verfahren ist vor allem darauf zurückzuführen, dass hier nicht notwendigerweise ein Trainingsdatensatz erstellt werden muss. Dennoch bietet es sich an, auch hier für die Validierung der Ergebnisse einen Teil der Daten händisch zu labeln. Auch die Komplexität der Verfahren kann eine Rolle dabei

spielen, welche Verfahren eingesetzt werden können. So sind regelbasierte Ansätze generell mit weniger technischem Know-how durchführbar, Machine und Deep-Learning Verfahren setzen hingegen tiefergehende technische und programmatische Expertise voraus.

Abbildung 3-5: Entwicklungsaufwand und Komplexität der verschiedenen Klassifikationsverfahren



Quelle: Eigene Darstellung

Zuletzt soll auch die Möglichkeit der Synthese verschiedener Ansätze diskutiert werden. Zunächst können unabhängig voneinander die verschiedenen Ansätze erprobt werden, um dann als Ergebnis jeweils einen Datensatz mit den relevanten Stellenanzeigen zu erzeugen. Allerdings kann es vorkommen, dass die Stellenausschreibungen bestimmter Berufe innerhalb der Automobil- und Zulieferindustrie besonders gut von regelbasierten Ansätzen und schlechter von Machine und Deep Learning Algorithmen erfasst werden können, für andere relevante Berufe dies jedoch umgekehrt gilt. In einem solchen Fall würde eine vereinheitlichte Ergebnisliste aus verschiedenen Verfahren die Fehler minimieren und ein möglichst gutes Ergebnis mit Blick auf die relevanten Stellenanzeigen liefern.

Ein zweiter übergreifender Ansatz kann in der Kombination der verschiedenen Verfahren bestehen. Sucht man etwa mit dem Unternehmenswörterbuch direkt nach den Stellenausschreibungen in Baden-Württemberg ansässiger Unternehmen aus der Automobil- und Zulieferindustrie, können hierdurch auch Stellenanzeigen aus nicht relevanten Berufsgruppen identifiziert werden (zum Beispiel, wenn Daimler nach einem Gartenlandschaftsbauer für seine Grünanlagen sucht). Identifiziert man eine genügend große Menge an Stellenanzeigen, können in einem nachfolgenden Schritt die beschriebenen Machine Learning Verfahren angewendet werden, um die Stellenanzeigen in für die jeweilige Branche relevante und nicht relevante Ausschreibungen zu unterscheiden.

Des Weiteren könnte das händisch erstellte und validierte Kompetenzwörterbuch (siehe Kapitel 3.2.2.1) verwendet werden, um die Annotation der Trainingsdaten für die NER-Modelle zu erleichtern. So kann das Wörterbuch im Trainingsdatensatz eine große Anzahl an Kompetenzen bereits erkennen und markieren. Menschliche Codierer müssen diese maschinellen Annotationen dann lediglich noch überprüfen und gegebenenfalls fehlende Kompetenzen markieren. Da die Annotation eines Trainingsdatensatzes für die Modellierung in NER-Verfahren meist sehr zeitaufwendig ist, kann die Kombination der beiden Verfahren eine deutliche Effizienzsteigerung bewirken.

3.2.2 Identifikation von Kompetenzen in Stellenausschreibungen der Automobil- und Zulieferindustrie

3.2.2.1 Kompetenzen mittels Wörterbuch-Ansatz

Um die relevanten Kompetenzen von Berufen der Automobil- und Zulieferindustrie aus den Stellenausschreibungen zu extrahieren, kann ähnlich wie in Kapitel 3.2.1.1 ein regelbasierter, iterativer Wörterbuchansatz genutzt werden. Im Unterschied zu dem oben diskutierten Ansatz sollen hier ausschließlich Kompetenzen in das Wörterbuch eingehen. Dafür werden die durch eine initiale Suchliste gefundenen Stellenanzeigen händisch nach zusätzlichen Kompetenzen durchsucht, die bisher nicht im Wörterbuch enthalten waren. Nach einem Validierungsprozess können die neu identifizierten Kompetenzen in das Wörterbuch mit aufgenommen werden und mit diesem neuen Wörterbuch kann eine neue Stichwortsuche in den Volltexten initiiert werden. Als Startpunkt für die Erstellung des Wörterbuchs könnte die in Kapitel 3.1 erarbeitete Kompetenzliste dienen.

Da die Erstellung und vor allem die Validierung von Wörterbüchern meist sehr zeitaufwendig sind, können verschiedene Text Mining Verfahren verwendet werden, um diesen Prozess zu beschleunigen. Hierfür können zum Beispiel Worthäufigkeitsmatrizen beziehungsweise tfidf-Matrizen der Stellenanzeigen erstellt werden. Kontrastiert man die identifizierten Stellenanzeigen der Automobil- und Zulieferindustrie mit Stellenanzeigen anderer Branchen, sollten die gewichteten tfidf Werte solcher Wörter höher gewichten, die (fast) ausschließlich in Ausschreibungen der Automobil- und Zulieferindustrie genutzt werden, nicht aber in anderen Ausschreibungen. Da diese Wörter zwischen Stellenanzeigen beziehungsweise Berufen differenzieren, sollten darunter auch Kompetenzen sein, die verstärkt in den jeweiligen Berufen der Automobil- und Zulieferindustrie nachgefragt werden. Die resultierenden Wortlisten können dann nach Gewichtung sortiert manuell überprüft werden. Handelt es sich um eine relevante Kompetenz, kann diese dem Wörterbuch samt möglicher Flexionsformen hinzugefügt werden.

3.2.2.2 KI-Verfahren für die Extrahierung relevanter Kompetenzen

Der oben beschriebene regelbasierte Ansatz mittels eines Wörterbuchs funktioniert meist gut, ermöglicht es jedoch nicht, auf Kompetenzen zu generalisieren, die nicht im Wörterbuch enthalten sind. Hierfür können jedoch Deep Learning Verfahren eingesetzt werden, wie etwa die auf neuronalen Netzen basierende „Named Entity Recognition“ (NER). Trainiert auf einen manuell annotierten Trainingsdatensatz kann ein solches neuronales Netz auch gänzlich neue

Kompetenzen identifizieren, da es den sprachlichen Kontext erkennt, in dem solche Kompetenzen aufgezählt oder beschrieben werden (siehe Li et al., 2020).

Für NER sind hochqualitative Trainingsdaten essenziell (Lample et al., 2016). Deshalb sollte für die Analyse ein Trainingsdatensatz definiert und manuell annotiert werden. Der Annotationsprozess umfasst die maschinenlesbare Markierung und kategoriale Zuordnung relevanter Wörter oder Wortkombinationen durch menschliche Codierer. Konkret bedeutet das, dass die in den Stellenanzeigen geforderten Kompetenzen in einer zufälligen Stichprobe händisch markiert werden. Für die Qualitätssicherung sollten hier unabhängig voneinander mehrere Codierer eingesetzt werden, die mit zuvor erstellten Annotations Guidelines arbeiten (siehe Benikova et al., 2014). Um eventuelle Konflikte zwischen den Markierungen zu beseitigen, überprüft üblicherweise eine weitere Person alle Annotationen. Da die Qualität des Trainingsdatensatzes sehr wichtig für den weiteren Analyseverlauf ist, sollten die Codierer entsprechende inhaltliche Expertise in geforderten Kompetenzen der Automobil- und Zulieferindustrie mitbringen. Zudem empfiehlt es sich schon an dieser Stelle, die so markierten Kompetenzen mit relevanten Stakeholdern abzusprechen.

Für die Annotation bieten sich mehrere Plattformen an, die Annotationstools produktionsfertig bereitstellen. Solche Tools ermöglichen die Markierung relevanter Wörter oder Wortkombinationen in Texten und die Transformation und den Export der annotierten Daten in gängige Formate, die von NLP- beziehungsweise NER-Frameworks verarbeitet werden können. Eine kostenpflichtige Plattform stellt etwa das Tool „prodigy“ (<https://prodi.gy/>) dar, eine gute Open Source Alternative bietet „Docco“, das sowohl auf einem lokalen Server als auch in einer Cloud Umgebung genutzt werden kann (Nakayama, 2018).

Generell kann NER als ein Teilgebiet der Informationsextraktion verstanden werden, welches das Ziel hat, benannte Objekte in unstrukturiertem Text zu identifizieren und in vordefinierte Entitäten beziehungsweise Klassen zu gruppieren (Awasthy et al., 2020). Dabei können benannte Objekte etwa die Namen von Personen, Orte, Organisationen oder eben auch Kompetenzen sein. Ein durch menschliche Codierer annotierter Beispieltext einer Stellenausschreibung für den Beruf eines „Data Scientist – automatisiertes Fahren (m/w/d)“ ist in Abbildung 3-6 dargestellt. Hier wurden neben relevanten Kompetenzen („Skill“), die sich vor allem auf Programmiersprachen (C, C++ und Python) sowie Kenntnisse in der Cloud Umgebung von Microsoft Azure auch weitere mögliche Klassen definiert, wie etwa der geforderte Abschluss, die dazugehörigen Fächer und die gewünschte Berufserfahrung. Die Klassen können je nach Bedarf definiert werden und sich auch ausschließlich auf geforderte Kompetenzen beschränken.

Abbildung 3-6: Annotierter Beispieltext über mehrere Klassen

Als Data Scientist - automatisiertes Fahren (m/w/d) bei Daimler sollten Sie **C SKILL** / **C++ SKILL** oder **Python SKILL** beherrschen und Erfahrung in der **Microsoft Azure Cloud SKILL** mitbringen. Darüber hinaus sollten Sie einen **Master Abschluss ABSCHLUSS** in **(Wirtschafts-)Informatik FÄCHER** oder **Mathematik FÄCHER** und **mindestens fünf Jahre Berufserfahrung BERUFSERFABUNG** in relevanten Berufen vorweisen können.

Quelle: Eigene Darstellung unter Verwendung von Doccano und spaCy

Um den oben angesprochenen Vorteil von NER-Modellen gegenüber regelbasierten Verfahren (die Möglichkeit, automatisiert relevante Kompetenzen zu finden, die nicht Teil des Wörterbuchs sind) zu verdeutlichen, zeigt Abbildung 3-7 die Vorhersage der Klassen für das in Abbildung 3-6 dargestellte Textbeispiel. Hier wurde ein von Büchel und Mertens (2020) trainiertes NER-Modell verwendet, um die Klassen im oben dargestellten Beispiel (siehe Abbildung 3-6) vorherzusagen.¹⁰ Zusätzlich zum dort dargestellten Beispieltext wurde die erfundene Programmiersprache „Fiktiv-X“ zum Text hinzugefügt. Diese fiktive Programmiersprache taucht im gesamten restlichen Textkorpus nicht auf und wurde dementsprechend nicht annotiert. Zudem würde sie auch in einem Wörterbuch nicht auftauchen. Abbildung 3-7 verdeutlicht, dass ein NER-Modell (siehe Büchel/Mertens, 2020) diese fiktive Programmiersprache korrekt als weitere Kompetenz identifizieren kann. Da NER-Modelle lernen, in welchem Kontext ein Wort benutzt wird, konnte darauf geschlossen werden, dass es sich beim Begriff „Fiktiv-X“ um eine Programmiersprache handeln muss. Das NER-Modell kann demnach gegenüber regelbasierten Ansätzen generalisieren, also auch gänzlich neue Begriffe identifizieren und der korrekten Klasse zuordnen.

Abbildung 3-7: Vorhersage eines NER-Modells für den Beispieltext

Als Data Scientist - automatisiertes Fahren (m/w/d) bei Daimler sollten Sie **C SKILL** / **C++ SKILL** , **Fiktiv-X SKILL** oder **Python SKILL** beherrschen und Erfahrung in der **Microsoft Azure Cloud SKILL** mitbringen. Darüber hinaus sollten Sie einen **Master Abschluss ABSCHLUSS** in **(Wirtschafts-)Informatik FÄCHER** oder **Mathematik FÄCHER** und **mindestens fünf Jahre Berufserfahrung BERUFSERFABUNG** in relevanten Berufen vorweisen können.

Quelle: Eigene Darstellung unter Nutzung des NER-Modells von Büchel / Mertens (2020)

Für die Implementierung von NER-Modellen zur Identifikation von Kompetenzen in Stellenausschreibungen werden hier beispielhaft zwei nützliche NLP-Frameworks vorgestellt, die in der Programmiersprache Python angeboten werden – Flair und spaCy. Flair ist ein state-of-the-art NLP Framework, das dem Nutzer erlaubt, verschiedene NLP Modelle wie NER, part-of-speech tagging (Pos) oder Textklassifizierung für eine stetig wachsende Anzahl verschiedener Sprachen zu

¹⁰ Büchel und Mertens (2020) nutzen das in Python implementierte NLP-Framework spaCy, um ein NER-Modell zu optimieren, welches verschiedene Entitäten in Stellenausschreibungen von KI-Berufen identifiziert.

implementieren (Akbik et al., 2019). Eine weitere verbreitete NLP-Bibliothek ist spaCy. Mit dem Release von Version 3.0 können ebenfalls state-of-the-art NLP-Modelle aufgesetzt werden, die ähnlich gute Ergebnisse liefern wie Flair (siehe Honnibal et al., 2020). Beide Frameworks sind Open Source Lösungen, die ohne zusätzliche Kosten lokal implementiert und optimiert werden können. Die Vorteile von spaCy belaufen sich vor allem auf die effiziente Implementierung. Das bedeutet, dass teilweise Algorithmen genutzt werden, die explizit die Laufzeit der jeweiligen Modelle verringern sollen. Da Deep Learning Modelle sehr rechenintensiv sind, kann es sich je nach vorhandener Hardware und Größe der verwendeten Datensätze anbieten, spaCy zu nutzen, da so in deutlich kürzerer Zeit gute Ergebnisse erzielt werden können. Im Gegensatz dazu kann das Trainieren von NER-Modellen in Flair deutlich mehr Rechenzeit in Anspruch nehmen. Flair liefert aber für viele NLP-Aufgaben bessere Ergebnisse. Ein Nachteil von Flair gegenüber spaCy ist die Dokumentation. Da Flair relativ neu ist, sind viele Funktionen nicht gut dokumentiert. Die Nutzungsweise von spaCy hingegen ist sehr ausführlich dokumentiert und bietet viele Beispiele, die bei der Implementierung der NER-Modelle für die eigene Nutzung hilfreich sein können.

Abseits dieser übergreifenden Frameworks besteht auch die Möglichkeit, neu entwickelte NER-Modelle direkt zu implementieren. Da der entsprechende Code fast immer Open Source zur Verfügung gestellt wird, kann dieser mit entsprechendem Know-how direkt übernommen und auf das eigene Problem angewendet werden.¹¹ Je nach technischer und programmatischer Expertise kann auf verschiedene Frameworks zurückgegriffen werden oder der Code direkt aus entsprechenden Open Source Repositorien übernommen werden.

In der Gesamtbetrachtung sind NER-Modelle am einfachsten in spaCy zu implementieren, gefolgt von Flair und der direkten Übernahme von Codes aus den Repositorien aktuell veröffentlichter Forschung. Betrachtet man die Performance verschiedener NER-Modelle und -Frameworks zeigt sich aber auch, dass aktuellere Modelle generell bessere Ergebnisse erzielen.¹²

3.2.2.3 Deskriptive Analyse der Kompetenzen

In einer deskriptiven Analyse sollten die so identifizierten Kompetenzen aufgearbeitet werden. Dieser Analyseschritt umfasst beispielsweise die Darstellung der wichtigsten beziehungsweise häufigsten Kompetenzen pro Berufsbezeichnung.

Dafür müssen die mit dem NER-Modell extrahierten Kompetenzen zunächst normalisiert werden. So kann etwa die Kompetenz der Erfahrung in Microsofts Cloud Plattform Azure mit verschiedenen Bezeichnungen wie „Azure“, „MS Azure“, „Microsoft Azure Cloud“, „Microsoft-Azure“, etc. umschrieben werden. Um die geforderten Kompetenzen über Berufe hinweg vergleichen zu können, müssen diese unterschiedlichen Schreib- und Ausdrucksformen in ein einheitliches Format (etwa „Microsoft Azure“) überführt werden. Je nach Anzahl der zu extrahierenden Klassen oder der in den Texten verwendeten Bezeichnungen können hier Wörterbuch-

¹¹ Für kürzlich veröffentlichte NER-Modelle mit entsprechendem Open Source Code siehe etwa Yamada et al. (2020), Wang et al. (2019) oder Straková et al. (2019).

¹² Für eine Übersicht über aktuelle NER-Ansätze siehe https://github.com/sebastianruder/NLP-progress/blob/master/english/named_entity_recognition.md.

oder Machine Learning Verfahren angewendet werden.¹³ Ist die Anzahl der extrahierten Begriffe eher gering, kann es effizienter sein, ein Wörterbuch zu erstellen, das die jeweiligen Begriffe vereinheitlicht. Bei einer sehr großen Anzahl an extrahierten Begriffen kann aber auch ein Klassifikator entwickelt werden, der entscheiden kann, ob zwei Begriffe synonym verwendet werden können oder nicht. Hierfür kann etwa ein Trainingsdatensatz erstellt werden, in welchem händisch jeweils die umschreibende und die einheitliche Klasse einander zugeordnet werden, und mit dessen Hilfe anschließend die übrigen Klassen automatisiert zugeordnet werden (siehe etwa Weston et al., 2019).

3.3 Zusammenführung der Arbeiten

Als Ergebnis der Arbeiten von Kapitel 3.1 und Kapitel 3.2 ergeben sich, unter anderem Listen von Kompetenzen je relevantem Beruf in der Automobil- und Zulieferindustrie. Diese Listen sind je Beruf für die weitere Analyse zusammenzuführen. In einem ersten Schritt sollte eine Normalisierung der Kompetenzlisten pro Beruf erfolgen (siehe beispielsweise Kapitel 3.2.2.3). Bei der Erstellung einheitlicher Kompetenzlisten aus verschiedenen Datenquellen (hier etwa die Kombination von Berufenet-Daten mit den Daten der Stellenausschreibungen) können verschiedene Situationen eintreten, die zu beachten sind (siehe auch Bertelsmann 2020, 15):

- (1) Einzelne Kompetenzen können mehrfach auftreten. Da die in Berufenet umschriebenen Kompetenzen auch in den Stellenanzeigen auftauchen können, sollten die Kompetenzen in einem nachfolgenden Schritt normalisiert beziehungsweise zusammengefasst werden.
- (2) Die Doppelungen können nicht nur aus denselben Begriffen mit leicht veränderter Schreibweise bestehen, sondern auch aus synonym verwendeten Begriffen. Hier ist gegebenenfalls ein tiefergehendes Verständnis der jeweiligen Kompetenzen notwendig, um diese sinnvoll zuordnen zu können. An dieser Stelle könnte sich eine weitere Einbindung der Stakeholder anbieten.
- (3) Zuletzt sind die verschiedenen hierarchischen Ebenen zu beachten, in denen Kompetenzen zusammengefasst werden können. So werden in Berufenet meist höhere Abstraktionsebenen verwendet als in Stellenanzeigen (etwa „Programmieren“ anstatt „R“ oder „Python“). Diese verschiedenen Eben müssen gegebenenfalls manuell zugeordnet werden, um Kompetenzen nur auf derselben Ebene miteinander zu vergleichen.

Je nach Anzahl der als relevant für die Automobil- und Zulieferindustrie identifizierten Berufe und der zugehörigen Kompetenzen kann dieser Prozess relativ zeitaufwendig sein, da die Kompetenzen überwiegend händisch einander zugeordnet werden müssen, um die Qualität des daraus resultierenden Datensatzes gewährleisten zu können. Methodische Expertise in der Datenverarbeitung und Fachexpertise im Bereich der Berufsbildung müssen für ein erfolgreiches Ergebnis eng miteinander verzahnt werden.

¹³ Technisch sind hier grundsätzlich die gleichen Herangehensweisen anzuwenden wie in Kapitel 3.2.2.1 und 3.2.1.3 beschrieben.

4 Ähnlichkeitsanalyse zwischen Berufen

Mit den so generierten Kompetenzlisten für jeden relevanten Beruf der Automobil- und Zulieferindustrie in Baden-Württemberg können zwischen den Berufen empirisch Ähnlichkeiten berechnet werden.

Durch die zuvor analysierten Arbeitsmarktindikatoren können Kompetenzen von Berufen mit steigender Nachfrage denen von Berufen mit abnehmender Nachfrage gegenübergestellt und verglichen werden. Mit diesen Informationen können im weiteren Prozess Ähnlichkeiten zwischen den Berufen mit fallender und steigender Nachfrage analysiert werden, um so darauf zu schließen, wie Übergänge gestaltet werden können.

In der relevanten Literatur gab es in den letzten Jahren eine Reihe an Publikationen, in welchen die Ähnlichkeiten zwischen verschiedenen Berufen berechnet wurden (siehe Kandera et al., 2020; Lamb et al., 2018; Mealy et al., 2018; World Economic Forum, 2018). Orientiert an Kandera et al. (2020) wird hier ein Verfahren beschrieben, mit welchem der größtmögliche Anteil an Informationen über die jeweiligen Kompetenzanforderungen eines Berufs genutzt wird, um die Ähnlichkeit zu anderen Berufen zu berechnen.¹⁴

(1) Generierung von Word Embeddings

In einem ersten Schritt werden für jede Kompetenz so genannte Word Embeddings generiert, welche die semantische Bedeutung einer Kompetenz in ihrem Kontext abbilden können.¹⁵ So kann nicht nur verglichen werden, ob dieselbe Kompetenz in verschiedenen Berufen gefordert wird, sondern auch, ob es andere Kompetenzen gibt, die inhaltlich ähnlich beschrieben werden.¹⁶ Um ebenfalls Informationen über den jeweiligen Kontext zu erfassen, können für Kompetenzen aus Stellenanzeige die Volltexte der Stellenanzeigen und für die Daten von Berufenet die jeweiligen Texte mit Berufsbeschreibungen genutzt werden.¹⁷

(2) Vergleich der Kompetenzen

Die jeweiligen Embeddings können über verschiedene statistische Methoden der Ähnlichkeitsberechnung zwischen Vektoren erhoben werden. Diese umfassen etwa die Kosinus-Ähnlichkeit,

¹⁴ Die Ergebnisse und die zugrundeliegenden Codes dieser Arbeit werden sukzessive über Github der Öffentlichkeit zur Verfügung gestellt (<https://github.com/nestauk/mapping-career-causeways>). Diese Vorarbeiten können als guter Startpunkt dienen, jedoch weichen der zu analysierende Datensatz, die Sprache und die Analyse substantiell von der hier angedachten Studienfrage ab. Daher ist nicht zu erwarten, dass die programmierten Codes übernommen werden können.

¹⁵ Für eine Diskussion über Word Embeddings siehe den Abschnitt über Deep Learning Verfahren in Kapitel 3.2.1.3.

¹⁶ Hier gilt es zu beachten, dass die Kombination der Kompetenzen aus verschiedenen Quellen (etwa Kompetenzen von Berufenet und aus Stellenanzeigen) methodisch unsauber sein könnte. Da sich die Word Embeddings dann jeweils auf verschiedene Textcorpora beziehen, könnten inhaltlich sehr ähnliche Kompetenzen einen niedrigen Ähnlichkeitswert bekommen, allein dadurch bedingt, dass die genutzte Sprache in den verschiedenen Quellen sehr unterschiedlich ist. Dies müsste während der Analyse empirisch überprüft werden und es müssten eventuell Lösungsmöglichkeiten entwickelt werden.

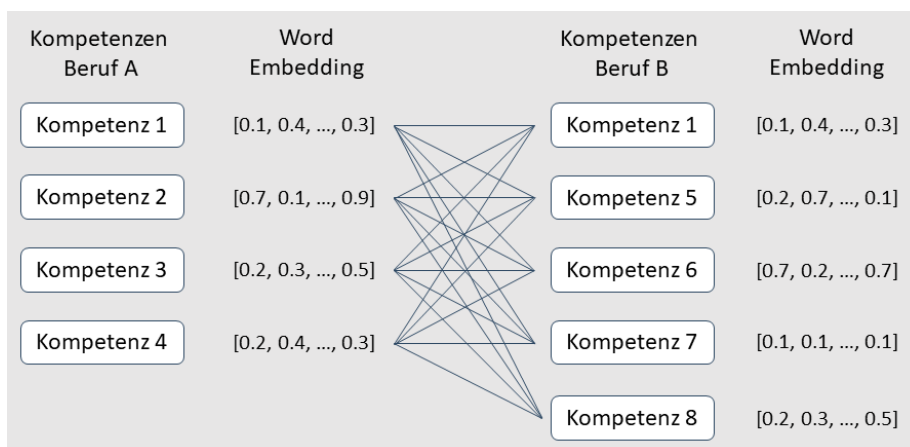
¹⁷ Auch für die Kompetenzen in ESCO stehen beschreibende Text zur Verfügung, aus welchen Word oder Sentence Embeddings generiert werden können (wie etwa bei Kandera et al., 2020).

die Euklidische Distanz, die durchschnittliche Manhattan Metrik oder die Word Mover's Distance. Üblicherweise wird hier die Kosinus-Ähnlichkeit berechnet, aber auch andere Metriken sollten empirisch überprüft werden, da diese, je nach Anwendungsfall, bessere Ergebnisse liefern können. Die Kosinus-Ähnlichkeit zwischen zwei Vektoren A und B (die Word Embeddings von zwei Kompetenzen), ist definiert über den Winkel (θ) zwischen den beiden Vektoren und lässt sich berechnen durch:

$$\text{Ähnlichkeit}(A, B) = \cos(\theta) = \frac{A * B}{||A|| ||B||}$$

Der resultierende Wert liegt beim Vergleich von Textdaten zwischen 0 (keine Ähnlichkeit) und +1 (identischer Text). Abbildung 4-1 zeigt beispielhaft die Auflistung von Kompetenzen zweier Berufe (Beruf A und Beruf B) und dessen Word Embedding Repräsentation. Hier wird dargestellt, dass für jede Kompetenz in beiden Berufen zunächst ein Word Embedding erstellt wird. Die daraus resultierenden Vektoren von Beruf A werden dann jeweils unter Verwendung der Kosinus-Ähnlichkeit mit allen anderen Vektoren von Beruf B verglichen.

Abbildung 4-1: Berechnung der Ähnlichkeiten zwischen Kompetenzen



Quelle: Eigene Darstellung

(3) Filtern und Zuordnung der Kompetenzen

Der nächste Schritt umfasst die Zuordnung aller Kompetenzen aus Beruf A mit der jeweils ähnlichsten Kompetenz aus Beruf B. Das bedeutet, dass die oben berechneten Ähnlichkeitswerte für Kompetenz 1 aus Beruf A mit allen Kompetenzen aus Beruf B (1, 5, 6, 7 und 8) zunächst sortiert werden und das Kompetenzpaar mit der höchsten Ähnlichkeit entsprechend ausgewählt wird. Dieser Prozess wird mit allen weiteren Kompetenzen aus Beruf A ebenfalls durchgeführt. Wichtig hierbei ist, dass die Kompetenzen aus dem Zielberuf (Beruf B) aus der Liste entfernt werden, sobald sie einmal zugeordnet worden sind. Werden in Beruf B mehr Kompetenzen gefordert als Beruf A bietet, wird eine Ähnlichkeit von 0.00 angegeben, da hier eine geforderte Kompetenz im Zielberuf nicht abgedeckt wird. Im umgekehrten Fall (Beruf A bietet mehr Kompetenzen als Beruf B) wird kein Wert zugeordnet und die zusätzliche Kompetenz aus Beruf A

taucht nicht in der Ähnlichkeitsberechnung auf. Abbildung 4-2 zeigt die beispielhafte Zuordnung der Kompetenzen der oben dargestellten Berufe. Hier sind die jeweils ähnlichsten Kompetenzpaare einander zugeordnet und die Ähnlichkeitswerte entsprechend angegeben. Da Kompetenz 1 identisch ist und in beiden Berufen gefordert wird, beträgt der Ähnlichkeitswert 1.00.

Abbildung 4-2: Berechnung der Ähnlichkeiten zwischen Berufen

Kompetenzen Beruf A	Kompetenzen Beruf B	Ähnlichkeit	Gewichtung
Kompetenz 1	Kompetenz 1	1.00	0.25
Kompetenz 3	Kompetenz 8	0.75	0.10
Kompetenz 2	Kompetenz 6	0.60	x 0.05 = 0.499
Kompetenz 4	Kompetenz 5	0.32	0.45
	Kompetenz 7	0.00	0.05

Quelle: Eigene Darstellung

(4) Berechnung der gemittelten Ähnlichkeitswerte

Der letzte Schritt umfasst dann üblicherweise die Mittelung der berechneten Ähnlichkeitswerte für jeden Beruf (siehe Kanders et al., 2020; Lamb et al., 2018). Hierfür bildet man die Summe aller Ähnlichkeitswerte und dividiert diese durch die Anzahl geforderter Kompetenzen in Beruf B. Wie oben beschrieben, ist dabei die Richtung des beruflichen Übergangs entscheidend und die resultierenden Werte sind dementsprechend asymmetrisch. Das bedeutet, dass der Übergang von Beruf A in Beruf B eine geringere Ähnlichkeit beschreibt als der Übergang von Beruf B in Beruf A.

Bei diesem Schritt vier ist jedoch kritisch anzumerken, dass keine Gewichtung der Relevanz, der Interaktion, der Transferierbarkeit oder der Komplexität der jeweiligen Kompetenzen vorgenommen wird. Für einen validen Vergleich verschiedener Berufe ist diese Einordnung aber von zentraler Bedeutung. Denn die gleiche Gewichtung aller Kompetenzen wie etwa bei Kanders et al. (2020) kann zu Verzerrungen in den aus den Ähnlichkeiten resultierenden Übergangspfaden führen. Nimmt man etwa an, dass die beispielhaften Kompetenzen 6, 7 und 8 von Beruf B wenig Bedeutung für die Berufsausübung haben oder relativ schnell erlernt werden können, die Kompetenzen 1 und vor allem 5 aber zentral sind und/oder nur unter hohem Zeitaufwand erlernt werden können, kann die jeweilige Einordnung einen großen Einfluss auf die tatsächliche Ähnlichkeit zwischen den Berufen haben. Deshalb empfiehlt es sich, wie in Abbildung 4-2 dargestellt, eine Gewichtung vorzunehmen, die die Bedeutung beziehungsweise die Komplexität der Kompetenzen des Zielberufs beschreibt. Der resultierende finale Ähnlichkeitswert kann dann

über das gewichtete arithmetische Mittel dargestellt werden.¹⁸ Für die Generierung dieser Gewichtungsvektoren ist ein tiefgehendes Verständnis der Berufe und Kompetenzen unerlässlich – Datenanalyse und eine Fachexpertise im Bereich der Berufsbildung sollten auch in diesem Analyseschritt Hand in Hand gehen. Idealerweise können hier zusätzlich die Stakeholder um eine Einschätzung der Relevanz der verschiedenen Kompetenzen gebeten werden, beispielsweise über eine Fragebogenmethode, bei der die Stakeholder die Bedeutung und/oder Komplexität der Kompetenzen auf einer festgelegten Skala einordnen.

Es ist zu überlegen, ob im weiteren Projektverlauf neben Automobilberufen auch weitere Berufe in der Ähnlichkeitsanalyse berücksichtigt werden sollten. Ist beispielsweise ein Ergebnis der Kompetenzanalyse, dass immer mehr IT-Kompetenzen von Bedeutung sind, sollten auch IT-Berufe mitberücksichtigt werden.

¹⁸ Das gewichtete arithmetische Mittel (\bar{x}) lässt sich berechnen durch $\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$, wobei x_i den jeweiligen Ähnlichkeitswert für die Kombination i an Kompetenzen darstellt und w_i die zugehörige Gewichtung. Sind die gewichteten Werte bereits normalisiert (w'_i), lässt sich das gewichtete arithmetische Mittel vereinfacht berechnen durch $\bar{x} = \sum_{i=1}^n w'_i x_i$.

5 Optionsmodul: Änderungen innerhalb von Berufen

Neben der Entwicklung von Übergangspfaden zwischen unterschiedlichen Berufen könnten ergänzend auch Entwicklungen innerhalb von Berufen berücksichtigt werden. Veränderungen von Kompetenzbedarfen innerhalb von Berufen können zum einen entstehen, wenn sich die berufsspezifischen Tätigkeiten stark verändern, sodass ältere Beschäftigte diese Kompetenzen in ihrer Ausbildung nicht erlernt haben. Entsprechend wird eine Person mit einer Ausbildung zum „Mechaniker für Karosserieinstandhaltungstechnik“ nach der alten Ausbildungsordnung nicht die gleichen Kompetenzen erworben haben wie ein Ausbildungsabsolvent der modernisierten Ausbildung zum „Kfz-Mechatroniker“ mit dem Schwerpunkt „Karosserietechnik“. Zum anderen ergeben sich die Änderungen, wenn in einem Beruf neue Tätigkeiten hinzukommen, die es früher in dieser Form nicht gab. So ist beispielsweise der Schwerpunkt „System und Hochvolttechnik“, der sich mit Antriebsarten wie Hybrid- und Elektro-Motoren beschäftigt, 2013 als völlig neuer Schwerpunkt in der Ausbildung des Kfz-Mechatronikers hinzugekommen. Bei den Kompetenzen, die neu zu den bestehenden Berufen hinzukommen, kann es sich um Kompetenzen handeln, die in anderen Berufen bereits gelernt und angewendet werden oder es kommen völlig neuartige Kompetenzen hinzu, die auch für andere Berufe neu sind.

Auch wenn sich die Tätigkeiten in Berufsgattungen im Zeitablauf verändern und damit ebenso die Kompetenzanforderungen an die Beschäftigten, bedeutet das nicht zwingend, dass Personen, deren Ausbildung schon länger zurück liegt, nicht über diese Kompetenzen verfügen. So können ältere Beschäftigte schon häufiger an Weiterbildungen teilgenommen und durch Berufserfahrung und Lernen im Prozess der Arbeit ihre Kompetenzen weiterentwickelt haben. Ein Indikator für Passungsprobleme könnte das gleichzeitige Vorhandensein vieler unbesetzter Stellen in einem Beruf bei gleichzeitig vielen (formal) passend qualifizierten Arbeitslosen sein. Um die Beschäftigungsfähigkeit möglichst vieler erhalten zu können, kann es daher notwendig sein, bestimmte Personengruppen gezielt weiterzubilden. Um aufzuzeigen, welche Kompetenzen innerhalb von Berufen aktuell vermehrt nachgefragt werden, ist es möglich, alte Stellenanzeigen mit neuen zu vergleichen. Ein Vergleich von älteren und neueren Ausbildungsordnungen bietet sich nur bedingt an, da sie sich weniger häufig ändern. Durch die technikneutrale und gestaltungsoffene Formulierung von Kompetenzen sind sie weniger konkret und Änderungsbedarfe werden erst in größeren Zeitabschnitten notwendig. Alte Fassungen von Berufen existieren nicht, so dass auch hier kein Zeitvergleich möglich ist. Wenn bestimmte Kompetenzen in den letzten Jahren vermehrt und andere Kompetenzen gar nicht mehr in Stellenausschreibungen genannt werden, könnte man daraus einen Trend sowie einen Qualifizierungsbedarf ablesen. Dabei wäre in Abstimmung mit den Stakeholdern zu klären, ob eine Fokussierung auf die Berufe mit einer steigenden Nachfrage als ausreichend erachtet wird, um den veränderten Kompetenzbedarf abzubilden.

6 Fazit

Die empirischen Arbeiten zur Abgrenzung der für die Automobil- und Zulieferindustrie relevanten Berufe, der dazugehörigen Kompetenzen sowie die Ähnlichkeitsanalysen sind zentrale Analyseschritte, um im weiteren Projektverlauf Übergangspfade zwischen Berufen mit sinkender und steigender Nachfrage in der Automobil- und Zulieferindustrie abzuleiten.

In dieser Expertise wurden vier vorbereitende Schritte zur Entwicklung von Übergangspfaden in der Automobil- und Zulieferindustrie vorgestellt und konkretisiert:

1. Identifikation von relevanten Berufen: In Schritt 1 sollen relevante Berufe in der Automobil- und Zuliefererindustrie identifiziert und in Berufe mit steigender und sinkender Nachfrage gruppiert werden. Wie in Kapitel 2 dargelegt, wird hierfür die Klassifikation der Berufe 2010 (KldB 2010) als grundlegende Taxonomie sowie die Verwendung Beschäftigtendaten der BA als relevante Datenquellen empfohlen.
2. Erstellung von Kompetenzprofilen: In Schritt 2 werden für die relevanten Berufe Kompetenzprofile erarbeitet. Kapitel 3.1 erläutert, inwiefern die Tätigkeitsbeschreibungen des Berufenet der BA hierfür eine gute Ausgangsbasis sind. Da Tätigkeits- und Kompetenzbeschreibungen häufig technologieneutral formuliert sind, werden als zweite Säule aktuelle Stellenanzeigen unter der Verwendung von Big Data Methoden ausgewertet, welche die nachgefragten Kompetenzen aus Unternehmenssicht konkretisieren (Kapitel 3.2). Aus der Synthese ergeben sich berufsspezifische Kompetenzprofile. Da die verschiedenen Quellen Kompetenzen in unterschiedlichen Abstraktionsniveaus darstellen, ist ein zentraler Arbeitsschritt bei der Synthese das Clustern von Kompetenzen.
3. Ähnlichkeitsanalyse zwischen Berufen: In Schritt 3 wird die Ähnlichkeit von Berufen basierend auf den Kompetenzprofilen analysiert. Dabei wird insbesondere überprüft, in welchen Kompetenzen sich Berufe mit wachsender beziehungsweise sinkender Nachfrage ähneln. Ein Ziel dieser Ähnlichkeitsanalyse ist, Vorarbeiten für die Entwicklung von Übergangspfaden von weniger gefragten oder gefährdeten Berufen in Berufe/Berufsgruppen mit wachsender Nachfrage zu entwickeln. Kapitel 4 beschreibt zum einen, wie unter der Verwendung von Big Data Methoden technisch bei der Ähnlichkeitsanalyse vorgegangen werden kann. Bei der Umsetzung ist für ein zielführendes Ergebnis neben technischem Know-how insbesondere Praxiswissen erforderlich, welches unter anderem die Bedeutung der jeweiligen Kompetenzen und auch deren Interaktion mit anderen erforderlichen Kompetenzen in den verschiedenen Berufen einordnet.
4. Änderungen innerhalb von Berufen (optional): Im vierten und letzten Schritt können ergänzend Änderungen innerhalb von Berufen analysiert werden. So können Qualifizierungspfade abgeleitet werden, die ältere Beschäftigte in einem Beruf in die Lage versetzen, aktuelle Anforderungen zu erfüllen. Hintergrund ist, dass die gestaltungsoffenen und technikneutralen Ausbildungsordnungen im dualen System viel Veränderung innerhalb von Berufen zulassen, sodass technologischer Wandel nicht automatisch zu neuen Berufen führt.

7 Literatur

Akbik, Alan / Bergmann, Tanja / Blythe, Duncan / Rasul, Kashif / Schweter, Stefan / Vollgraf, Roland, 2019, FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP, in: Proceedings of the NAACL-HLT 2019: Demonstrations, S. 54–59

Ali, Jehad / Khan, Rehanullah / Ahmad, Nasir / Maqsood, Imran, 2012, Random Forests and Decision Trees, in: IJCSI International Journal of Computer Science, 9. Jg., Nr. 5, S. 272–278

Amato, Flora / Boselli, Roberto / Cesarini, Mirko / Mercorio, Fabio / Mezzanzanica, Mario / Moscato, Vincenzo / Persia, Fabio / Picariello, Antoni, 2015, Classification of Web Job Advertisements: A Case Study, in: 23rd Italian Symposium on Advanced Database Systems (SEBD 2015), Gaeta, Italien, S. 144–151

Awasthy, Parul / Moon, Taesun / Ni, Jian / Florian, Radu, 2020, Cascaded Models for Better Fine-Grained Named Entity Recognition, arXiv preprint, <https://arxiv.org/abs/2009.07317> [26.11.2020]

Benikova, Darina / Biemann, Chris / Rezniek, Marc, 2014, NoSta-D Named Entity Annotation for German: Guidelines and Dataset, in: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), S. 2524–2531

Blei, David M. / Ng, Andrew Y. / Jordan, Michael, I., 2003, Latent Dirichlet Allocation, in: Journal of Machine Learning Research, 3. Jg., S. 993–1022

Boselli, Roberto / Cesarini, Mirko / Marrara, Stefania / Mercorio, Fabio / Mezzanzanica, Mario / Pasi, Gabriella / Viviani, Marco, 2018, WoLMIS: a labor market intelligence system for classifying web job vacancies, in: Journal of Intelligent Information Systems, 51. Jg., S. 477–502

Breiman, Leo, 2001, Random Forests, in: Machine Learning, 45. Jg., Nr. 1, S. 5–32

Büchel, Jan / Mertens, Armin, 2020, Nachfrage nach KI-relevanten Qualifikationen, BMWi Publikation, im Erscheinen

Bundesagentur für Arbeit – BA, 2020, Grundlagen: Qualitätsbericht – Statistik der sozialversicherungspflichtigen und geringfügigen Beschäftigung, Nürnberg, November 2020, <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Methodik-Qualitaet/Qualitaetsberichte/Generische-Publikationen/Qualitaetsbericht-Statistik-Beschaeftigung.pdf?blob=publicationFile&v=7> [24.11.2020]

BA, 2019, Meldeverfahren zur Sozialversicherung – Schlüsselverzeichnis für die Angaben zur Tätigkeit, https://www.arbeitsagentur.de/datei/dok_ba015567.pdf [20.11.2020]

BA, 2017, Betriebsnummernvergabe, https://www.arbeitsagentur.de/datei/dok_ba015200.pdf, [24.11.2020]

BA, 2011a, Klassifikation der Berufe 2010 – Band 1: Systematischer und alphabetischer Teil mit Erläuterungen, https://www.arbeitsagentur.de/datei/Klassifikation-der-Berufe_ba017989.pdf [20.11.2020]

BA, 2011b, Umsteigeschlüssel von der KldB 2010 zur ISCO-08, <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010/Arbeitshilfen/Umsteigeschluessel/Generische-Publikation/Umsteigeschluessel-KldB2010-ISCO-08.xls?blob=publicationFile&v=5> [20.11.2020]

Burstedde, Alexander / Flake, Regina / Jansen, Anika / Malin, Lydia / Risius, Paula / Seyda, Susanne / Schirner, Sebastian / Werner, Dirk, 2020, Die Messung des Fachkräftemangels – Methodik und Ergebnisse aus der IW-Fachkräftedatenbank zur Bestimmung von Engpassberufen und zur Berechnung von Fachkräftelücken und anderer Indikatoren, IW-Report 59/2020, <https://www.iwkoeln.de/studien/iw-reports/beitrag/alexander-burstedde-regina-flake-anika-jansen-lydia-malin-paula-risius-susanne-seyda-sebastian-schirner-dirk-werner-die-messung-des-fachkraeftemangels.html>, [23.11.2020]

Cajner, Tomaz / Ratner, David, 2016, A Cautionary Note on the Help Wanted Online Data, FEDS Notes, Board of Governors of the Federal Reserve System, <https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/a-cautionary-note-on-the-help-wanted-online-data-20160623.html> [20.11.2020]

Colombo, Emilio / Mercurio, Fabio / Mezzanzanica, Mario, 2018, Applying machine learning tools on web vacancies for labour market and skill analysis, https://techpolicyinstitute.org/wp-content/uploads/2018/02/Colombo_paper.pdf [23.11.2020]

Demary, Vera / Engels, Barbara / Goecke, Henry / Koppel, Oliver / Armin Mertens / Christian Rusche / Marc Scheufen / Jan Wendt, 2020, KI-Monitor 2020. Status quo der Künstlichen Intelligenz in Deutschland, Gutachten im Auftrag des Bundesverbands Digitale Wirtschaft (BVDW) e.V.

Devlin, Jacob / Chang, Ming-Wei / Lee, Kenton / Toutanova, Kristina, 2019, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1. Jg., S. 4171–4186

Europäische Kommission, 2020, ESCO – Europäische Klassifikation für Fähigkeiten, Kompetenzen, Qualifikationen und Berufe, <https://ec.europa.eu/esco/portal/occupation> [01.12.2020]

Hemnes, Jürgen / Schandock, Manuel, 2016, Stellenanzeigenanalyse in der Qualifikationsentwicklungsforschung. Die Nutzung maschineller Lernverfahren zur Klassifikation von Textabschnitten, Bundesinstitut für Berufsbildung, Bonn, <https://www.bibb.de/veroeffentlichungen/de/publication/show/8146> [14.12.2020]

Hershbein, Brad / Kahn, Lisa B., 2018, Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings, in: American Economic Review, 108. Jg., Nr. 7, S. 1737–1772

Honnibal, Matthew / Montani, Ines / Van Lendeghem, Sofie / Boyd, Adriane, 2020, Introducing spaCy v3.0 nightly, <https://explosion.ai/blog/spacy-v3-nightly> [14.12.2020]

Kanders, Karlis / Djumalieva, Jyldyz / Sleeman, Cath / Orlik, Jack, 2020, Mapping Career Causeways: Supporting workers at risk, London: Nesta, <https://www.nesta.org.uk/report/mapping-career-causeways-supporting-workers-risk/> [3.12.2020]

Jurafsky, Daniel / Martin, James H., 2019, Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, <https://web.stanford.edu/~jurafsky/slp3/> [26.11.2020]

Lamb, Creig / Huynh, Annalise / Vu, Viet, 2019, Lost and Found: Pathways from Disruption to Employment, Toronto: Brookfield Institute, <https://brookfieldinstitute.ca/lost-and-found-pathways-from-disruption-to-employment/> [3.12.2020]

Lample, Guillaume / Ballesteros, Miguel / Subramanian, Sandeep / Kawakmi, Kazuya / Dyer, Chris, 2016, Neural Architectures for Named Entity Recognition, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, S. 260–270

Li, Jing / Sun, Aixin / Han, Jianglei / Li, Chenliang, 2020, A Survey on Deep Learning for Named Entity Recognition, in: IEEE Transactions on Knowledge and Data Engineering, S. 1–20

McCallum, Andrew / Nigam, Kamal, 1998, A comparison of event models for Naïve Bayes text classification, in: Proceedings of the AAAI-98 Workshop on Learning for Text Categorization, S. 41–48

Mealy, Penny / del Rio-Chanona, R. Maria / Farmer, J. Dooyne, 2018, What you do at work matters: New lenses on labour, SSRN Electronic Journal, doi.org/10.2139/ssrn.3143064 [3.12.2020]

Minaee, Shervin / Kalchbrenner, Nal / Cambria, Erik / Nikzad, Narjes / Chenaghlu, Meysam / Gao, Jianfeng, 2020, Deep Learning Based Text Classification: A Comprehensive Review, <https://arxiv.org/abs/2004.03705> [23.11.2020]

Mimno, David / Wallach, Hanna M. / Talley, Edmund / Leenders, Miriam / McCallum, Andrew, 2011, Optimizing semantic coherence in topic models, in: EMNLP '11: Proceedings of the Conference on Empirical Methods in Natural Language Processing, S. 262–272

Nakayama, Hiroki / Kubo, Takahiro / Kamura, Junya / Taniguchi, Yasufumi / Liang, Xu, 2018, doccano: Text Annotation Tool for Human, <https://github.com/doccano/doccano> [26.11.2020]

O’Kane, Layla / Narasimhan, Rohit / Nania, Julia / Taska, Bledi, 2020, Digitalization in the German Labor Market. Analyzing Demand for Digital Skills in Job Vacancies, Bertelsmann Stiftung, Gütersloh, <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/digitalization-in-the-german-labor-market-en> [14.12.2020]

Rennie, Jason D. M. / Shih, Lawrence / Teevan, Jaime / Karger, David R., 2003, Tackling the Poor Assumptions of Naïve Bayes Text Classifiers, in: Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), S. 616–623

Roberts, Margaret E. / Stewart, Brandon M. / Tingley, Dustin / Airoidi, Edoardo M., 2013, The Structural Topic Model and Applied Social Science, in: Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application and Evaluation, <https://scholar.princeton.edu/bstewart/publications/structural-topic-model-and-applied-social-science> [26.11.2020]

Rohrbach-Schmidt, Daniela / Hall, Anja, 2020, BIBB/BAuA-Erwerbstätigenbefragung 2018, Daten und Methodenbericht, Heft 1/2020, <https://www.bibb.de/veroeffentlichungen/de/publication/show/16401> [26.11.2020]

Statistisches Bundesamt, 2008, Klassifikation der Wirtschaftszweige – Mit Erläuterungen, <https://www.destatis.de/static/DE/dokumente/klassifikation-wz-2008-3100100089004.pdf> [20.11.2020]

Stops, Michael / Bächmann Ann-Christin / Glassner, Ralf / Janser, Markus / Matthes, Britta / Metzger, Lina-Jeanette / Müller, Christoph / Seitz, Joachim, 2020, Machbarkeitsstudie Kompetenz-Kompass. Teilprojekt 2: Beobachtung von Kompetenzerfordernissen in Stellenangeboten, <https://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/Forschungsberichte/fb-553-machbarkeitsstudie-kompetenz-kompass.pdf?blob=publicationFile&v=1> [20.11.2020]

Straková, Jana / Straka, Milan / Hajic, Jan, 2019, Neural Architectures for Nested NER through Linearization, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, S. 5326–5331

Tong, Simon / Koller, Daphne, 2001, Support Vector Machine Active Learning with Applications to Text Classification, in: Journal of Machine Learning Research, in: Journal of Machine Learning Research, 2. Jg., Nr. 1, S. 45–66

Turrell, Arthur / Speigner, Bradley J. / Djumalieva, Jyldyz / Copple, David / Thurgood, James, 2019, Transforming Naturally Occurring Text Data into Economic Statistics: The Case of Online Job Vacancy Postings, NBER Working Paper No. 25837, May 2019

Wang, Zihan / Shang, Jingbo / Liu, Liyuan / Lihao, Lu / Liu, Jiacheng / Han, Jiawei, 2019, Cross-Weigh: Training Named Entity Tagger from Imperfect Annotations, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), S. 5154–5163

Weston, Leigh / Tshitoyan, Vahe / Dagdelen, John / Kononova, Olga / Persson, Kristin / Ceder, Gerbrand / Jain, Anubhav, 2019, Named Entity Recognition and Normalization Applied to Large-Scale Information Extraction from the Materials Science Literature, in: Journal of Chemical Information and Modeling, 59. Jg, Nr. 9, S. 3692–3702

World Economic Forum, 2018, Towards a Reskilling Revolution: A Future of Jobs for All, [weforum.org/docs/WER_FOW_Reskilling_Revolution.pdf](https://www.weforum.org/docs/WER_FOW_Reskilling_Revolution.pdf) [3.12.2020]

Yamada, Ikuya / Asai, Akari / Shindo, Hiroyuki / Takeda, Hideaki / Matsumoto, Yuji, 2020, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), S. 6442–6454

Yang, Zhilin / Dai, Zihang / Yiming, Yang / Carbonell, Jaime / Salakhutdinov, Ruslan / Le, Quoc V., 2019, XLNet: Generalized Autoregressive Pretraining for Language Understanding, in: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada

Tabellenverzeichnis

Tabelle 2-1: Anforderungsniveaus in der KldB 2010	6
Tabelle 2-2: Vergleich KldB 2010 und ISCO-08	10
Tabelle 2-3: Maschinenbauingenieure in der ISCO-08 und der KldB 2010	10
Tabelle 2-4: Ausgewählte Engpass-Indikatoren in den drei Kraftfahrzeugberufen für Baden-Württemberg.....	12

Abbildungsverzeichnis

Abbildung 2-1: Die Automobil- und Zulieferindustrie in der Klassifikation der Wirtschaftszweige (WZ 2008).....	7
Abbildung 2-2: Veränderung der Beschäftigung in den drei Kraftfahrzeugtechnik-Berufen in Baden-Württemberg (Normiert 2013 = 100)	12
Abbildung 3-1: Beispielhafte Online-Stellenanzeige für einen „Inbetriebnehmer (m/w/d) Kraftfahrzeugtechnik“	18
Abbildung 3-2: Iterativer Prozess der Identifikation von Stellenanzeigen über eine Wörterbuchsuche mit Berufsbezeichnungen.....	20
Abbildung 3-3: Beispielhafte Häufigkeitsmatrix für alle Wörter im Gesamtkorpus aller Stellenanzeigen.....	22
Abbildung 3-4: Beispielhafte tfidf-Matrix für alle Wörter im Gesamtkorpus der Stellenanzeigen.....	23
Abbildung 3-5: Entwicklungsaufwand und Komplexität der verschiedenen Klassifikationsverfahren.....	26
Abbildung 3-6: Annotierter Beispieltext über mehrere Klassen	29
Abbildung 3-7: Vorhersage eines NER-Modells für den Beispieltext.....	29
Abbildung 4-1: Berechnung der Ähnlichkeiten zwischen Kompetenzen	33
Abbildung 4-2: Berechnung der Ähnlichkeiten zwischen Berufen.....	34