

# **Kapitel 5 – Repräsentationen der Datenwissenschaften im schweizerischen Arbeitsmarkt**

---

## **5.1 Einleitung**

»Sherlock Holmes wäre von [Ihrem] Data-Spürsinn begeistert? Ist für [Sie] Data Science der rote Faden[,] der zum Erfolg führt? Sie haben einen Hochschulabschluss in Wirtschaftsinformatik, Informatik, Statistik, Mathematik oder Physik oder eine vergleichbare Ausbildung. Sie haben in ein oder mehreren Technologien und Analysetools ihre Erfahrungen gesammelt. Interessant sind zum Beispiel Technologien wie Python, R, Scala, Hadoop, Spark, SAP HANA ... Oder Analysetools wie Hue, SAS Customer Intelligence, SPSS, SAP BO, Matlab, Shiny, Splunk, TensorFlow, Tableau, Spark und Qlikview? ... Zögern [Sie] nicht, mir Ihre Fähigkeiten mitzuteilen! Neben Ihrer Data-Kompetenz liegen Ihre Stärken in geschicktem, initiativem, zielorientiertem, verantwortungsbewusste[m] und unternehmerischem Denken, Handeln und Verhandeln. Sie können gesprächspartnerorientiert komplexe Sachverhalte kompetent, einfach und verständlich kommunizieren. Auch fungieren Sie gerne selbstständig und sind zugleich ein intuitiver Teamplayer. Sie schätzen es[,] Ihre Umgebung für Ihre innovativen Ideen und Projekte zu begeistern!«<sup>1</sup>

Das einleitende Zitat beschreibt das Profil einer »Persönlichkeit« für eine Position als Data Scientist im Schweizer Arbeitsmarkt. Auf der inhaltlichen Ebene wird deutlich, dass eine Person mit quantitativ-technischem Fachhintergrund gesucht wird, die mit verschiedenen Programmiersprachen und Datenanalysetools versiert ist und zudem unternehmerische, soziale und kommunikative Kompetenzen aufweist – ein »Skillset«, das so oder ähnlich für verschiedene Berufsbezeichnungen artikuliert wird, gleichzeitig allerdings äusserst selten vorzufinden ist. Trotzdem steht es exemplarisch für zeitgenössische technikwissenschaftliche Wissensgebiete (Paulitz & Prietl 2017: 158ff.).

---

<sup>1</sup> Stelleninserat »Data Scientist: Customer Journey« des Unternehmens Addexpert GmbH: <https://www.addexpert.ch/jobs/job-detail.aspx?Bereich=Informatik&JobID=6244> (Original nicht mehr verfügbar; letzter Zugriff: 04.07.2017, Ergänzungen zur besseren Lesbarkeit in eckigen Klammern durch den Autor).

Interessant an diesem Ausschnitt ist, dass sich das Aufzählen äquivalenter Alternativen durch die verschiedenen Ebenen des skizzierten Persönlichkeitsprofils zieht. Das Zitat beginnt gleich mit zwei Metaphern, die unterschiedliche Deutungsangebote für die Persönlichkeit als Data Scientist bereitstellen: Zum einen wird das Bild einer detektivisch-analytischen Persönlichkeit (»Sherlock Holmes«) bemüht, die dank ihrem »Data-Spürsinn« komplexe Fälle lösen kann. Es wird also eine persönliche Eigenschaft der möglichen Bewerber\*innen angesprochen, die eine führende (literarische) Referenzfigur der kriminalistischen Forensik »begeister[n]« würde, sofern sie denn existierte. Ein zweites Bild wird durch das Grundmotiv des »rote[n] Faden[s]« evoziert, wonach in »Data Science« die leitende Strukturierung besteht, »[die] zum Erfolg führt«. Die zweite Metapher setzt »Data Science« also als Mittel zum Erfolg, wobei unklar im Inserat insgesamt bleibt, um welche Art von Erfolg (kommerziell, wissenschaftlich, in der Auflösung von Kriminalfällen etc.) es sich handelt.

Auf die zwei unterschiedlichen Imaginationen zur Persönlichkeit von Data Scientists zum Einstieg folgt eine Aufzählung von disziplinären (»Wirtschaftsinformatik, Informatik, Statistik, Mathematik oder Physik«) und qualifikatorischen (»Hochschulabschluss [...] oder vergleichbare Ausbildung«) Hintergründen. Diese werden in der Stellenanzeige als Grundlage des Persönlichkeitsprofils präsentiert, ohne dies allerdings explizit zu machen. Kryptisch wird es anschliessend für Nicht-Insider\*innen, wenn die gesuchten »Erfahrungen« und »Fähigkeiten« in knapp zwei Dutzend unterschiedlichen »Technologien« und »Analysetools« ausbuchstabiert werden. Schliesslich nennt der Ausschnitt eine ganze Reihe von Kompetenzen, d. h. individuell zugeschriebenen Eigenschaften, die sehr unterschiedliche Betätigungsfelder und Praktiken indizieren, jedoch divergente Entstehungshintergründe aufweisen (Salvisberg 2010).

Strukturell, so lässt sich der zitierte Ausschnitt des Stelleninserats interpretieren, bleibt also unklar, aus welchen Imaginationen, Disziplinen, Qualifikationen, Technologien, Analysetools und Kompetenzen sich die gesuchte Persönlichkeit von Data Scientists zusammensetzt. Sie präsentiert sich als eine Aufzählung vermeintlich äquivalenter Zuschreibungen und Eigenschaften. Sinngemäß sind solche Aufzählungen im Sinne »unverbindliche[r] Wunschlisten« (Geser 1983: 479) ein konstitutives Merkmal von Stellenanzeigen, um ein möglichst breites Spektrum an Bewerber\*innen anzusprechen. Der Forschungsstand hat aber deutlich gemacht, dass sich eine Vielzahl unterschiedlicher Akteur\*innen mit der Frage nach den Begriffen, Kompetenzen und Persönlichkeitsprofilen der Datenwissenschaften beschäftigt. Die Suche nach und das Zusammenstellen der »richtigen« »Skillsets«, die sich im zitierten Ausschnitt des Stelleninserats exemplarisch manifestieren, kulminierte in den letzten Jahren zu einem zentralen Erkenntnisinteresse im Untersuchungsfeld selbst. Ein Merkmal der Diskussion besteht darin, umfangreiche Listen unterschiedlicher Methoden, Tools, Bildungsqualifikationen oder individueller Fähigkeiten und Eigenschaften zu artikulieren. Das Formulieren solcher äquivalenter Alternativen signalisiert insofern sowohl Inklusivität bezüglich disziplinärer Wissensbestände, Qualifikationen oder methodisch-technischer Expertisen als auch Unklarheit darüber, was noch dazugehört und was nicht.

Das Kapitel fragt danach, wie solche Bedeutungszuschreibungen über die Datenwissenschaften durch unterschiedliche Listen und Aufzählungen multipler Begrifflichkeiten konstruiert werden. Allgemein erfuhren die Datenwissenschaften im Arbeitsmarkt in den letzten Jahren einen signifikanten Bedeutungsgewinn, der sich

in einer steigenden Anzahl von Stellenangeboten für Data Scientists äussert. Synchron dazu werden die Datenwissenschaften in der gesellschaftlichen Diskussion vornehmlich mittels ökonomischer Deutungen, wie der Diagnose eines gravierenden Fachkräftemangels (Manyika et al. 2011; Markow et al. 2017), gerahmt (Saner 2019). Wie ich noch ausführen werde, hat dies auch Effekte auf andere Felder: So werden die finanzielle Förderung von Forschungsinitiativen durch die Politik sowie die rasche Implementation neuer Studiengänge an Hochschulen und Universitäten primär mit der hohen Nachfrage nach entsprechenden Abschlüssen und Fähigkeiten im Arbeitsmarkt begründet. Stellenanzeigen für Data Scientists stellen insofern eine wichtige Konstruktionsleistung der Datenwissenschaften durch Akteur\*innen in verschiedenen Feldern dar. Die Analyse ihrer strukturellen und inhaltlichen Charakteristika ist deshalb elementar für das Verständnis der Datenwissenschaften als Wissenschaften.

Wie die Analyse zeigen wird, sind es neben wissenschaftlichen Institutionen vor allem Unternehmen in ökonomischen Feldern, aber auch öffentliche Einrichtungen, Non-Profit- sowie internationale Organisationen, die Kompetenzprofile von und für Data Scientists konstruieren. Sie alle bauen »Analytics Teams« auf (J. G. Harris et al. 2013; Kim et al. 2016; Patil 2011), formulieren Tätigkeitsbeschreibungen, publizieren Stelleninserate und rekrutieren schliesslich Datenwissenschaftler\*innen. In den Anforderungen an Stellenbewerber\*innen verdichten sich einerseits intraorganisationale Rollenerwartungen (Geser 1983), andererseits werden feld- und organisationsspezifische Perspektiven darauf formuliert, worin datenwissenschaftliche Praktiken und Expertisen bestehen, wo die Berührungs punkte sowie Grenzziehungen zu verwandten Gebieten liegen. Stellenanzeigen als Stellungnahmen kollektiver Akteur\*innen repräsentieren die Wahrnehmung des Wissensfeldes nach aussen genauso wie innerhalb der jeweiligen Organisation, d. h., sie fixieren durch sprachliche Kategorien eine bestimmte Deutung der Datenwissenschaften und privilegieren diese gegenüber anderen möglichen Bedeutungen. Das Changieren zwischen und das Ringen um die relevanten Kategorien, das sich in Stellenanzeigen in umfangreichen Listen kontingenter Werkzeuge, Methoden oder Bildungsanforderungen manifestiert, strukturiert umgekehrt die Erwartungshaltungen von Akteur\*innen in unterschiedlichen Feldern. Indem deren multiple, divergierende Praktiken dadurch koordiniert werden, fundiert dies die Datenwissenschaften als zwischenräumliches Phänomen und trägt zu dessen Stabilisierung bei.

Stellenanzeigen für Data Scientists und verwandte Berufsbezeichnungen ermöglichen es nun, von einem makroskopischen Standpunkt aus die Konstruktionsleistungen unterschiedlicher Akteur\*innen zu analysieren, aber auch Gemeinsamkeiten und Unterschiede der involvierten Felder herauszuarbeiten. Das Kapitel erhebt keinen Anspruch auf eine umfassende soziologische Analyse des schweizerischen Arbeitsmarktes (dazu Aratnam 2012; Kriesi et al. 2010; Sacchi et al. 2005; Salvisberg 2010), sondern es fragt nach den für das Segment der Datenwissenschaften zentralen Organisationen, Feldern, Profilen und Bedeutungszuschreibungen. Folgende Fragestellungen leiten die Analyse an:

- Wie ist das Segment der Datenwissenschaften im Arbeitsmarkt strukturiert? Welche Organisationen in welchen Feldern publizieren Stelleninserate für Data Scientists?
- Welche Profile von Data Scientists werden in den Inseraten konstruiert?

- Wie werden die Datenwissenschaften im Arbeitsmarkt repräsentiert, d. h., welche Deutungen werden in den Stelleninseraten entworfen?

Die Datengrundlage zur Untersuchung der Fragestellungen bilden Stelleninserate für Data Scientists und verwandte Professionen, die ich mittels Webscraping-Verfahren auf einer Schweizer Online-Jobplattform erhoben habe (dazu Kap. 4.2). Das resultierende umfangreiche Korpus an Textdokumenten habe ich mithilfe von Topic Modeling, einem computerlinguistischen Verfahren, ausgewertet.

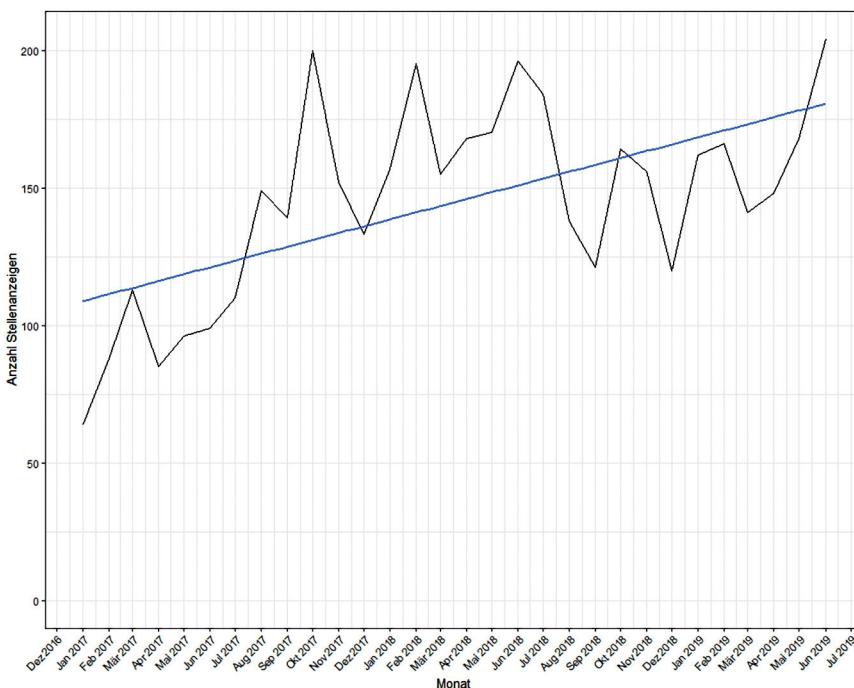
Das Kapitel ist wie folgt strukturiert: Zuerst erläutere ich anhand der deskriptiven Auswertung der Daten die Entwicklung, untersuche die zentralen Begriffe sowie die feldspezifische, geographische und sprachliche Verteilung der Datenwissenschaften im schweizerischen Arbeitsmarkt (Kap. 5.2). Den Hauptteil des Kapitels bilden die Analysen der Repräsentationen der Datenwissenschaften in den englischsprachigen (Kap. 5.3) sowie den deutschsprachigen Stellenanzeigen (Kap. 5.4). Schliesslich diskutiere ich die zentralen Ergebnisse im Hinblick auf die übergeordneten Fragestellungen der Arbeit (Kap. 5.5).

## 5.2 Entwicklung und Struktur des Arbeitsmarktsegments Datenwissenschaften

### 5.2.1 Entwicklung

Die Anzahl Stellenanzeigen nimmt über den Erhebungszeitraum hinweg betrachtet leicht, aber kontinuierlich zu. Sind die Stellenanzeigen nach Monaten gruppiert, wird die Entwicklung visuell nachvollziehbar (vgl. Abbildung 3). Durchschnittlich sind pro betrachteten Monat rund 145 Anzeigen erfasst. Die meisten Anzeigen wurden im Juni 2019 publiziert ( $n = 204$ ), die wenigsten im Januar 2017 ( $n = 66$ ). Zudem sind gewisse Schwankungen nach Jahreszeiten zu erkennen, beispielsweise die rückläufigen Zahlen zum Jahresende (Dezember 2017 und 2018), während in den darauffolgenden Monaten ein Anstieg zu beobachten ist. Dies ist möglicherweise die Folge regulatorischer Beschränkungen: Arbeitsbewilligungen für sogenannte »Drittstaatenangehörige«, d. h. Arbeitskräfte ausserhalb der EU/EFTA-Staaten, die primär für hochspezialisierte Berufstätigkeiten vergeben werden, wie es die untersuchten Professionen darstellen, sind stark kontingentiert (Aratnam 2012). Dafür spricht der hohe Anteil englischsprachiger Stellenanzeigen, der als ein Indikator für deren internationale Reichweite konzipiert werden kann, um ein breites Feld an Kandidat\*innen anzusprechen, das nicht auf nationalstaatliche bzw. sprachräumliche Grenzen limitiert ist.

Abbildung 3: Entwicklung der Stellenanzeigen im Korpus

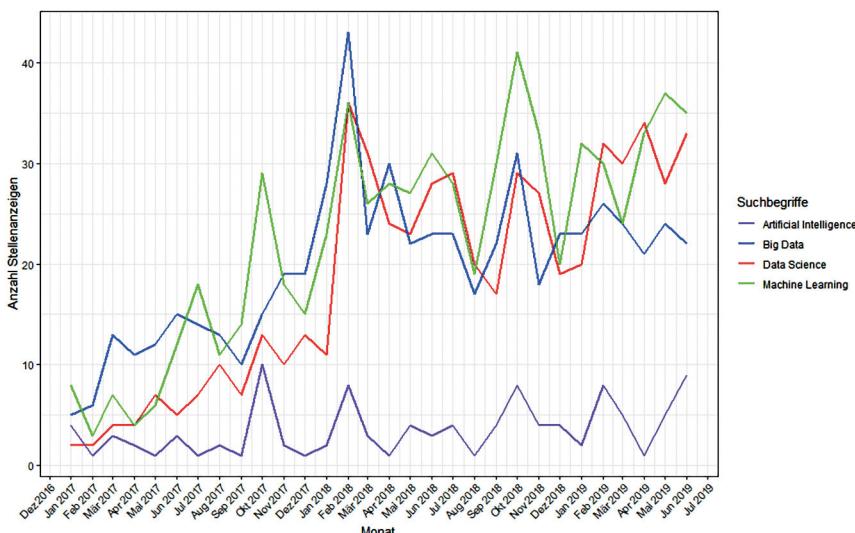


Bei der Unterscheidung der Stelleninserate nach den Begriffen Data Science, Big Data, Machine Learning und Artificial Intelligence offenbaren sich nur leicht divergierende Entwicklungsverläufe (vgl. Abbildung 4): Trotz inhaltlicher Differenzen weisen die drei Kategorien »Data Science«, »Big Data« und »Machine Learning« ähnliche Entwicklungen wie das Gesamttotal der Stellenanzeigen auf.<sup>2</sup> Obwohl »Machine Learning« nicht zur Erhebung des Samples verwendet wurde, bildet der Begriff die Entwicklung sogar leicht besser ab als »Data Science« und erfährt im Oktober 2018 einen Peak. Dies verweist darauf, dass »Machine Learning« zwar als Synonym bzw. Äquivalent für »Data Science« verwendet wird, während umgekehrt »Data Science« einen umfassenderen Begriff darstellt. Die Anzeigen für den Suchbegriff »Big Data« erreichen nach starkem Wachstum im ersten Erhebungsjahr im Februar 2018 einen Höchststand und sind danach rückläufig. Trotz seiner gegenwärtigen Konjunktur ist der Begriff »Artificial Intelligence« nur selten in den Stellenanzeigen vertreten.<sup>3</sup>

2 Die Korrelation der drei Datenreihen im Verhältnis zur Gesamtentwicklung ist mit Pearson R-Werten von 0.68 (»Big Data«), 0.75 (»Data Science«) und 0.81 (»Machine Learning«) hoch. Der Korrelationskoeffizient Pearson R ist ein Mass für die Stärke und Richtung eines linearen Zusammenhangs zwischen zwei intervallskalierten Variablen X und Y (Diaz-Bone & Weisger 2015: 230).

3 »Artificial Intelligence« bildet zudem die Gesamtentwicklung am schlechtesten ab (Pearson R = .53).

Abbildung 4: Entwicklung der Stellenanzeigen nach Suchbegriffen



Die temporale Entwicklung verläuft ähnlich, wenn die Stellenanzeigen nach Berufstiteln, Feldern oder geographischen Regionen differenziert betrachtet werden. Insgesamt erfährt das Segment der Datenwissenschaften demnach eine kontinuierlich zunehmende Relevanz im schweizerischen Arbeitsmarkt.

### 5.2.2 Häufige Begriffe in den Inseratetexten

Die Texte der Stelleninserate bilden das Hauptuntersuchungsobjekt des Kapitels. Mittels einfacher Wortfrequenzen und Vergleichen lassen sich deskriptiv Muster erkennen, noch bevor komplexere Verfahren wie Topic Modeling zur Anwendung kommen.

Im englischsprachigen Korpus wird – wenig erstaunlich – *data* (n = 21'479) am meisten verwendet. Weitere zentrale Terme sind *experience*, *team*, *work*, *skills*, *knowledge*, *working*, *job*, *support* oder *strong*, die charakteristisch für die Textgattung Stellenanzeigen sind. Spezifischer für das Sample sind Begriffe wie *business*, *development*, *management*, *research*, *science*, *analytics*, *analysis* oder *software*, die bestimmte Tätigkeitsfelder indizieren. Des Weiteren sind auch Rollenbezeichnungen (*scientist*, *analyst*, *engineer* etc.) sowie Standorte (*switzerland*, *basel*, *zürich*) unter den hundert häufigsten Termen zu finden.

Auch im deutschsprachigen Sample ist *data* (n = 3'582) der wichtigste Begriff, der immerhin in 92.6 % der Anzeigen erscheint, oft in Kombination mit weiteren englischsprachigen Termen *analytics*, *big* oder *science*. Dies zeigt, dass die Verwendung kategorialer Komposita wie »Data Science« oder »Big Data« sich auch im deutschen Sprachraum durchgesetzt hat. Begriffe wie Datenwissenschaften oder das gelegentlich verwendete Massendaten (Kolany-Raiser et al. 2018: 76ff.) finden sich nicht in den deutschsprachigen Inseraten. *daten* (n = 781) als deutschsprachiges Pendant zum Suchbegriff ist jedoch immerhin in 60.9 % der deutschsprachigen Dokumente vertreten. Weiter zeigen sich beim allgemeinen Vokabular diverse Überschneidungen (*erfah-*

rung, bereich, kenntnisse, team, arbeiten, suchen, aufgaben etc.) zum englischen Korpus, aber auch einige sprachspezifische Unterschiede wie *gute*, *kunden*, *m* (= »männlich«), *w* (= »weiblich«), *unsere*\* oder *ag*.

Inhaltlich interessant ist, dass sprachübergreifend verschiedene Analysetools bzw. Programmierumgebungen (*sql*, *r*, *python*, *excel*, *sap*) unter den hundert häufigsten Begriffen vertreten sind. Dies verweist einerseits auf die fundierende Bedeutung solcher Tools und Devices für die Datenwissenschaften. Andererseits werden Methoden und Analysetools wie erwähnt in den Inseratetexten oft austauschbar und in nicht-hierarchisierenden Aufzählungen angeführt (vgl. Beispelinserat 1). Da methodische Praktiken in einer Vielzahl von Softwareumgebungen programmierbar sind, rückt die Bedeutung des konkreten Tools in den Hintergrund – so können auch in SPSS »neuere Ansätze des [...] Machine Learnings« implementiert werden (Beispelinserat 1). Stellenanzeigen für Data Scientists verwenden demnach ein breites, inklusives Verständnis der erforderlichen methodischen und analytischen Tools.<sup>4</sup> Als *boundary objects* erfüllen sie die multiplen Anforderungen der Praxisfelder und können den jeweiligen lokalen Bedingungen angepasst werden, behalten dennoch ihre Identität und Stabilität über Feldgrenzen hinweg.

### 5.2.3 Berufsgruppen

Über die Titel der Stellenanzeigen lassen sich verschiedene Gruppen bilden, die auf unterschiedlichen Rollenbezeichnungen basieren. Die Einteilung in verschiedene Berufsgruppen dient dazu, um spezifische Zuschreibungen von Fertigkeiten, technologischen Kenntnissen oder Bildungsanforderungen im Sample zu untersuchen (Wowczko 2015). Darüber lassen sich auch die verschiedenen Profile von Data Scientists erschließen, die in der Literatur angeführt werden (vgl. Kap. 3.2). Besonders relevant sind die Gruppen der *Scientists* (23.3 %), der *Analysts* (17.6 %) sowie der *Engineers* (13.7 %). Die drei Gruppen stehen demnach für 54.6 % aller Anzeigen. Ebenfalls von Bedeutung sind die Gruppen *Manager* (7.8 %), *Specialists* (6.4 %), *Interns* (6.1 %), *Consultants* (5.1 %) sowie *Developer* (4.8 %). Die zehn wichtigsten Berufsgruppen im Sample machen rund 90 % des Samples aus (vgl. Tabelle 9 im Anhang). Betrachtet man die exakten Stellenbezeichnungen, so machen die drei zentralen Stellenbezeichnungen im Sample – *Data Scientist* (n = 587), *Data Analyst* (n = 497) und *Data Engineer* (n = 128) – rund 29 % der Anzeigen aus (vgl. Tabelle 10 im Anhang). Für den Berufstitel *Data Scientist* existiert keine deutsche oder französische Entsprechung im Sample – ein Hinweis darauf, dass sich die Rollenbezeichnung auch für anderssprachige Stelleninserate im Arbeitsmarkt etabliert hat.

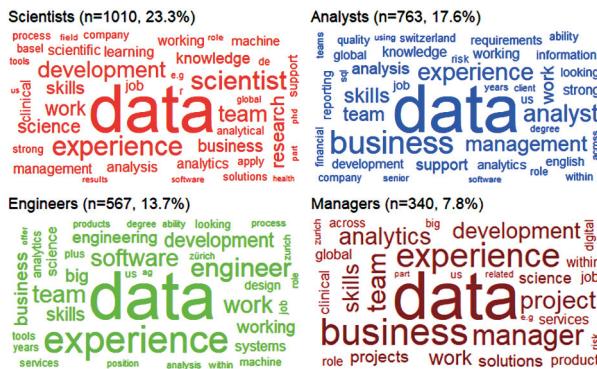
Über die Titel lassen sich zudem oft auftauchende Attribute eruieren (vgl. Abbildung 12 im Anhang): In rund 24 % der Anzeigen werden Pensenangaben verwendet. Am häufigsten sind 100 % (12.6 %), gefolgt von 80–100 % (5.8 %) sowie 80 % (3.9 %). Rund 18 % der Anzeigen enthalten geschlechtsindizierende Angaben (wie *male/female* oder *männlich/weiblich* bzw. deren Abkürzungen). Terme wie *Senior* (12.8 %), *Junior* (3.4 %), *Lead* (1.9 %) oder *Associate* (1.8 %) können der beruflichen Hierarchie der ausgeschriebenen Stellen zugeordnet werden. Auch andere Begriffe sind inhaltlich identifizierbar:

<sup>4</sup> Hier offenbart sich ein Widerspruch zu curricularen Konzeptionen an Universitäten und Hochschulen, die datenwissenschaftliche Expertise explizit in der »Fähigkeit zu skripten und programmieren« verorten und »nicht bloss in SPSS das richtige Menü auszuwählen« (Prof\_UH\_A).

Diese sind einerseits in konkreten Berufsbezeichnungen enthalten (wie *Business Analyst*), indizieren andererseits aber auch ein bestimmtes Tätigkeitsfeld, wie beispielsweise *Business* (7.6 %), *Software* (5 %), *Science* (4.6 %), *Research* (3.5 %) oder *Management* (3.2 %). Attribute wie *digital* (2.9 %), *clinical* (2 %) oder *technical* (1.4 %) charakterisieren gewisse Tätigkeitsfelder (*digital marketing* oder *clinical research*) oder Berufsbezeichnungen (*technical consultant*). In einzelnen Fällen werden auch Programmiersprachen oder Softwarelösungen (wie *SAP*, *R* oder *Java*) erwähnt.

Die Analyse häufiger Begriffe nach aggregierten Berufsgruppen fördert einige Unterschiede zutage (vgl. Abbildung 5): In der Gruppe der *Scientists* indizieren Terme wie *development, science, research, clinical, learning* oder *scientific* wissenschaftliche Tätigkeitsfelder. Das pharmazeutische Feld (*pharmaceutical* und *basel*) ist prominent in dieser Berufsgruppe vertreten. Bei den Tools figuriert die Statistiksprache *R* unter den dreissig häufigsten Begriffen. Bei den *Analysts* überwiegen demgegenüber für das ökonomische Feld generische Begriffe wie *business, management, reporting, quality* oder *services*. Trotzdem verweisen *risk, zürich* oder *financial* auf die primäre Verortung im Banken- bzw. im Versicherungsfeld. Unter den Tools findet *sql* am meisten Erwähnung. Bei den *Engineers* wiederum sind es vor allem *software, big, systems, design* oder *technology*. Tokens wie *zürich, swisscom, computer* und *platforms* geben Hinweise auf entsprechende Organisationen des informations- bzw. kommunikationstechnologischen Feldes. Hier ist *python* als einziges Tool unter den fünfzig häufigsten Begriffen vertreten.

Abbildung 5: Wordclouds der häufigsten Begriffe nach ausgewählten Berufsgruppen



Die Differenzierung der untersuchten Inserate über die Stellentitel legt unterschiedliche Profile von datenbezogenen Rollenbezeichnungen wie *Data Scientists*, *Data Analysts*, *Data Engineers* und weiteren offen. Wie bereits gezeigt prägte die Suche und Identifizierung entsprechender Profile von Beginn weg die Genese der Datenwissenschaften als soziales Phänomen. Die Ergebnisse der Analyse deuten signifikante Unterschiede hinsichtlich feldspezifischer Verortung, Standorten, Arbeitsinhalten, Qualifikationsanforderungen oder methodischen Vorgehensweisen an. Die identifizierten Begriffskonvolute repräsentieren das untersuchte Arbeitsmarktsegment insofern als ein heterogenes Wissensgebiet, das multiple epistemische Praktiken, disziplinäre Wissensbestände sowie technologische Objekte umfasst, die sich – wie ich im folgenden Unterkapitel zeigen werde – nicht auf einzelne soziale Felder begrenzen lassen.

## 5.2.4 Organisationen und Felder

Über die Organisationen, welche die Stellen ausschreiben, können die ökonomischen und sozialen Felder erschlossen werden, in denen Datenwissenschaftler\*innen gesucht werden und tätig sind.<sup>5</sup> Das Sample enthält 908 Nennungen unterschiedlicher Organisationen, die achtzehn unterschiedlichen Feldern entstammen. Dazu zählen vierzehn ökonomische Felder, das universitäre Feld, das Feld der staatlichen Verwaltung, das Feld der internationalen und Nichtregierungsorganisationen (NGO) sowie das hybride Feld der Gesundheit (Brown & Barnett 2004).<sup>6</sup> Bei der Analyse der Häufigkeiten fällt sofort die schiefe Verteilung der Inserate ins Auge: So verantworten zehn Organisationen einen Viertel aller Anzeigen (vgl. Tabelle 3), 65 Organisationen stehen bereits für rund 50 % aller Inserate. Demgegenüber gibt es 161 Organisationen, die zweimal im Sample enthalten sind, und 408 Organisationen mit nur einem Inserat. Das Pharmaunternehmen F. Hoffmann-La Roche ist mit 331 Inseraten (7.6 %) am häufigsten im Sample vertreten. Es folgen die Grossbanken Credit Suisse mit 125 Inseraten (2.9 %) und UBS mit 118 Anzeigen (2.7 %), das Telekommunikationsunternehmen Swisscom (2.5 %) sowie das Pharmaunternehmen Novartis (2.4 %).

Die Aufschlüsselung der Inserate nach einzelnen Feldern verdeutlicht, dass die fünf Organisationen mit den meisten Stellenanzeigen im Pharma-, Telekom/IKT- sowie im Bankenfeld zu verorten sind. Es handelt sich um Unternehmen, die alle ihren Hauptsitz in der Schweiz haben. Des Weiteren zählen der weltweit grösste Hersteller von Tabakprodukten (Philip Morris), die ETH Zürich, zwei international tätige Firmen der Personalrekrutierung (Harvey Nash und Experis) sowie Nestlé zu den zehn am häufigsten inserierenden Organisationen. Pharma und Banken sind auch über alle Stellenanzeigen betrachtet jene Felder, die am meisten Stelleninserate für Data Scientists publizieren: Das pharmazeutische Feld alleine verantwortet 19 % aller Anzeigen, während das Feld der Banken, Finanzdienstleister sowie FinTech-Unternehmen kombiniert 14.3 % des Samples ausmachen. Damit vereinen die zwei Felder bereits einen Drittelpart aller Anzeigen. Wichtige Felder sind auch das IT-/Software-Feld, das akademische Feld, das Feld der Consultingunternehmen, die Industrie, Versicherungen, Analytics-Firmen und Start-ups, Transport und Logistik, Unternehmen der Informations- und Kommunikationstechnologie (IKT) sowie der Detailhandel (mit jeweils über 4 %). Trotz der Dominanz einiger weniger Grossunternehmen indiziert die Verteilung insgesamt dennoch eine breite Nachfrage nach datenwissenschaftlicher Expertise über die achtzehn unterschiedlichen Felder hinweg. Dies deutet an, dass sich das Phänomen Datenwissenschaften keineswegs nur auf technologieintensive, datengetriebene Felder konzentriert, sondern im Sinne einer Datengesellschaft (Houben & Prietl 2018) auch soziale Bereiche affiziert, die bis dato als nur geringfügig datafiziert wahrgenommen werden, d. h. deren Funktionsweisen und Entscheidungsverfahren sich elementar auf die Erhebung, Prozessierung und Auswertung von Daten stützen.

5 Die folgenden Angaben beziehen sich auf die jeweiligen Organisationen inkl. Tochterunternehmen.

6 Das Feld der Gesundheit umfasst private, halbstaatliche und staatliche Akteur\*innen, die sich zwischen Gewinnorientierung, Gemeinwohlorientierung und Mischformen differenzieren lassen (Brown & Barnett 2004).

Tabelle 3: Die zehn am häufigsten inserierenden Organisationen im Sample

Organisation	Anzahl	Anteil	Feld
F. Hoffmann-La Roche AG	331	7.6 %	Pharma
Credit Suisse AG	125	2.9 %	Banken/FinTech
UBS AG	118	2.7 %	Banken/FinTech
Swisscom AG	108	2.5 %	Telekom/IKT
Novartis AG	105	2.4 %	Pharma
Philip Morris	77	1.8 %	Tabak
ETH Zürich	69	1.6 %	Wissenschaft/Bildung
Harvey Nash AG	62	1.4 %	Personalrekrutierung
Experis	59	1.4 %	Personalrekrutierung
Nestlé	41	0.9 %	Lebensmittel
<b>Total Sample</b>	<b>4341</b>	<b>100 %</b>	

Interessant ist zudem die Aufschlüsselung der identifizierten Berufsgruppen nach einzelnen Feldern: Das pharmazeutische Feld schreibt mit deutlichem Abstand die meisten Stellen in der Berufsgruppe *Scientist* aus (41.3 %), gefolgt von der Tabak (31.4 %) und der Lebensmittelindustrie (28.4 %). Die Stellenbezeichnung *Data Scientist* wird am meisten von Organisationen in den Feldern Versicherungen (24.8 %), Detailhandel (24.4 %) sowie Transport und Logistik (23 %) verwendet. Es handelt sich oft um Organisationen, die hochgradig datafiziert sind (Muster & Büchner 2018). *Data Science* taucht am häufigsten in den Stellenausschreibungen der öffentlichen Verwaltung (17.4 %), Universitäten und Hochschulen (11.3 %) auf, während sie in den anderen Feldern kaum vorkommt. Die staatliche Verwaltung sowie das akademische Feld greifen demnach auf die Kategorie im Sinne eines epistemischen Feldes zurück, während in den ökonomischen Feldern in erster Linie die Rollenbezeichnung *Data Scientist* dominiert.

Die Gruppe der *Analysts* macht im Banken/FinTech-Feld (34.7 %) über einen Drittel aller Anzeigen aus, gefolgt von den Versicherungen (31.6 %), Medien und Kommunikation (31.34 %) sowie dem Detailhandel (29.7 %). *Engineers* werden am häufigsten im Telekom/IKT-Feld (34.1 %), bei Analytics-Firmen und Start-ups (25.3 %), der Industrie (19.2 %) sowie in Software-Unternehmen (12.4 %) gesucht. Bei den übrigen Berufsgruppen fällt insbesondere der relativ hohe Anteil an *Consultants* bei Beratungsfirmen

(31.2 %) sowie im IT-Feld (17.1 %) auf. *Internships* und *Praktika* in Datenwissenschaften werden oft in der Industrie (14.4 %), im Detailhandel (7.7 %) sowie in der Pharmabranche (6.3 %) vergeben. Schliesslich sind die hohen Anteile anderer Berufsbezeichnungen an Universitäten und Hochschulen (47.2 %) sowie in der Pharmabranche (17.3 %) bemerkenswert. Eine genauere Inspektion zeigt, dass an Universitäten und Hochschulen feldspezifische Stellenbezeichnungen wie *Postdocs* (20.2 %), *Assistent\*innen* (10.7 %), *Mitarbeiter\*innen* (5.2 %), *Researcher* (5 %), *Dozent\*innen* (4.5 %) sowie *Professor\*innen* (3.9 %) oft vorkommen. Pharmaunternehmen suchen *Lead/er* (5.6 %), *Directors* (3.8 %), *Heads* (3.4 %) sowie ebenfalls *Postdocs* (3 %), die in den anderen Feldern selten sind.

## 5.2.5 Geographische Verteilung

Das untersuchte Arbeitsmarktssegment ist grossmehrheitlich auf wenige urbane Zentren in der Deutschschweiz konzentriert: Die Stadt Zürich ist (31.2 %) der häufigste Standort von Organisationen, die Stelleninserate für Data Scientists publizieren. An zweiter Stelle folgt Basel (16.1 %), was wie gezeigt am Pharmafeld liegt, in dem besonders viele Stellen ausgeschrieben werden. Die Städte Zürich und Basel vereinen also bereits knapp die Hälfte (47 %) aller Inserate und sind somit in Relation zur Bevölkerung signifikant übervertreten. Es folgen mit deutlichem Abstand die Städte Genf (6.3 %), Lausanne (5.2 %) sowie Bern (4.9 %). Zu den zehn häufigsten Standorten zählen ferner Winterthur und Luzern (je 1.6 %), Zug (1.5 %), Neuchâtel (1.4 %) sowie Dübendorf (1.1 %). Die zehn wichtigsten Standorte stehen für rund 71 % der Inserate im Sample.

Die Standorte illustrieren – wie die Sprachverteilung – die Fokussierung auf die deutschsprachige Schweiz gegenüber den anderen Sprachregionen. Betrachtet man die Grossregionen gemäss BFS, so verstärkt sich die Konzentration auf die Agglomerationen Zürich und Basel noch zusätzlich: 40.3 % aller Inserate werden im Kanton Zürich ausgeschrieben, der mit der Grossregion Zürich identisch ist.<sup>7</sup> Weitere 21.1 % der Inserate finden sich in der Grossregion Nordwestschweiz, zu der die Kantone Basel-Stadt, Baselland und Aargau zählen. Die beiden Grossregionen bilden zusammen über 61 % aller Anzeigen im Sample ab. Es folgt die Genferseeregion, zu der die Kantone Genf, Waadt und Wallis gezählt werden, mit 15.6 %. Die zweisprachige und einwohnerstärkste Grossregion Espace Mittelland (mit den Kantonen Bern, Solothurn, Freiburg, Neuchâtel und Jura) umfasst lediglich 10.6 %. Ebenfalls weniger bedeutend sind die Grossregionen Zentralschweiz (5.5 %), Ostschweiz (3.1 %) sowie das Tessin (0.7 %).<sup>8</sup> Es manifestiert sich in der Analyse eine ausgeprägte Konzentration auf wenige urbane Zentren, die zudem über ökonomische, politische sowie kulturelle Organisationen in ein globales Feld (Buchholz 2016) von Grossstädten eingebunden sind (Sassen 2013, 2016). Daraus lässt sich schliessen, dass Organisationen und Unternehmen in peripheren Regionen noch weniger datafiziert sind und auf grössere Schwierigkeiten bei der Besetzung von offenen Stellen stossen (Stadelmann 2017).

<sup>7</sup> Auch ohne die Stadt Zürich weist die Grossregion Zürich beinahe gleich viele Inserate (9.1 %) auf wie die drei Grossregionen Zentralschweiz, Ostschweiz und das Tessin zusammen.

<sup>8</sup> 2 % der Anzeigen enthalten Mehrfachnennungen (wie Zürich oder Basel), 0.3 % der Anzeigen sind im Ausland angesiedelt und 0.9 % liessen sich keinem eindeutigen Standort zuordnen. Für diese Inserate konnte folglich keine Kodierung nach Grossregion vorgenommen werden.

Tabelle 4: Die zehn häufigsten Standorte im Sample

Standort	Anzahl	Anteil	Feld mit höchstem Anteil
Zürich	1355	31.2 %	Banken/FinTech
Basel	700	16.1 %	Pharma
Genève	272	6.3 %	Transport/Logistik
Lausanne	225	5.2 %	Tabak
Bern	214	4.9 %	Transport/Logistik
Winterthur	71	1.6 %	Wissenschaft/Bildung
Luzern	68	1.6 %	Versicherungen
Zug	67	1.5 %	Pharma
Neuchâtel	61	1.4 %	Tabak
Dübendorf	47	1.1 %	Wissenschaft/Bildung
Total Sample	4341	100 %	

In der geographischen Verteilung spiegelt sich auch die Konzentration der jeweiligen Felder an bestimmten Standorten: In Zürich weist das Bankenfeld mit 32.8 % den höchsten Anteil<sup>9</sup> auf, wobei diverse weitere Felder ebenfalls häufig vertreten sind (wie IT/Software, Consulting, Analytics/Start-ups, Wissenschaft oder Versicherungen, je 5–11 %). In Basel hingegen vereint das pharmazeutische Feld fast drei Viertel aller Inserate (74 %), während andere Felder nur selten auftauchen. Dahingegen weisen Städte wie Genf, Lausanne und Bern eine heterogenere Verteilung auf mit verschiedenen Feldern, die mit Anteilen zwischen 10 und 20 % im Sample vertreten sind. In Luzern (Versicherungen), Zug (Pharma) oder Neuchâtel (Tabak) wiederum dominieren ebenfalls einzelne ökonomische Felder.<sup>10</sup>

9 Die folgenden Werte umfassen keine Inserate von Personalrekrutierungsfirmen.

10 In der Betrachtung nach Grossregionen schwächt sich diese Verteilung ab: Lediglich in der Nordwestschweiz dominiert das Feld der Pharma mit einem Anteil von knapp zwei Dritteln (65.5 %). Zu erwähnen ist ferner die Ostschweiz, wo der Anteil an Industriefirmen (34.6 %) deutlich höher liegt als derjenige anderer Felder. In den übrigen Grossregionen findet sich eine gleichmässigere Verteilung über die verschiedenen Felder hinweg.

## 5.2.6 Sprachverteilung

Sprachlich dominieren Stelleninserate auf Englisch: Insgesamt sind zwei Drittel aller Anzeigen (n = 2941, 67.7 %) auf Englisch verfasst, rund 30 % (n = 1282, 29.5 %) auf Deutsch. Obwohl auch französisch- (n = 117, 2.7 %) und italienischsprachige (n = 1, 0.02 %) Anzeigen enthalten sind, bildet das Sample wie erwähnt primär die deutschsprachige Schweiz ab.<sup>11</sup> Dabei ist zu beobachten, dass der Titel bzw. die Stellenbezeichnung und der Text der Stellenanzeige oft nicht dieselbe Sprache aufweisen. So hat ein signifikanter Teil der deutschsprachigen Anzeigen einen englischen Titel. Von den drei wichtigsten Rollenbezeichnungen sind jeweils rund zwei Drittel auf Englisch verfasst (vgl. Tabelle 10 im Anhang). Beispielsweise enthält das Sample 497 Anzeigen für *Data Analysts*. Davon sind wiederum 59 % englisch, 34.4 % auf Deutsch und 6.4 % auf Französisch verfasst. Demgegenüber gibt es lediglich 38 Anzeigen für *Datenanalyst\*innen* auf Deutsch sowie 11 für *Analyste de données* auf Französisch. Für *Data Scientist* finden sich gar keine deutsch- bzw. französischsprachigen Äquivalente im Sample.<sup>12</sup> Die Ergebnisse deuten darauf hin, dass sich die englischsprachige Professionsbezeichnung demnach bereits vollständig etabliert hat. Trotzdem sind ein Drittel der Anzeigen unter dieser Professionsrolle auf Deutsch oder Französisch verfasst.<sup>13</sup>

Die Sprachverteilung gibt schliesslich auch Hinweise auf die Internationalisierung der untersuchten Felder: So verfassen Unternehmen der Tabakindustrie (98.8 %) und jene der Pharmabranche (97 %) Stelleninserate fast ausschliesslich auf Englisch. Ebenfalls überdurchschnittlich hohe Anteile weisen internationale Organisationen und NGOs (83.6 %), das Lebensmittefeld (83.2 %) sowie das akademische Feld (79.3 %) auf. Die englischsprachigen Anteile sind also – neben dem akademischen Feld, das in der Schweiz stark internationalisiert ist – primär in jenen ökonomischen Feldern überdurchschnittlich, die von der Präsenz einiger weniger internationaler Grosskonzerne geprägt sind. Es ist insofern naheliegend, dass diese ihre Rekrutierungsaktivitäten stärker als andere Organisationen auf ein international mobiles Segment des Arbeitsmarktes ausrichten. Demgegenüber ist der Anteil englischsprachiger Inserate im Detailhandel (54.4 %), im Gesundheitswesen (48.6 %), im Telekom/IKT-Feld (46.5 %), im Transport- und Logistikfeld (46.1 %), im Energie- und Baufeld (31.9 %) sowie bei Versicherungen (27.4 %) unterdurchschnittlich.<sup>14</sup> Den Gegenpol markiert schliesslich das Feld der öffentlichen Verwaltung, wo keine Stelle auf Englisch verfasst ist. Auch hier

<sup>11</sup> Dies dürfte in erster Linie auf den Umstand zurückzuführen sein, dass eine Deutschschweizer Stellenplattform die Grundlage für das Webscraping darstellt (vgl. Kap. 4.2.2).

<sup>12</sup> Überdies verwendet lediglich ein Inserat die deutschsprachige Bezeichnung »Datenwissenschaften« im Text.

<sup>13</sup> Besonders hohe Anteile englischsprachiger Anzeigen sind bei den Berufsgruppen der Experts (86 %), der Scientists (77.9 %) und bei den Interns/Praktikant\*innen (76 %) zu beobachten (vgl. Tabelle 9). Umgekehrt ist der Anteil englischsprachiger Anzeigen bei den Gruppen der Consultants (39 %), der Specialists (54.9 %) sowie der Analysts (60.5 %) unterdurchschnittlich. Bei Letzteren deuten die hohen Anteile deutsch- und französischsprachiger Inserate an, dass Arbeitskräfte gesucht werden, die eine der Landessprachen beherrschen.

<sup>14</sup> Insofern ist es erstaunlich, dass die nur wenig internationalisierten Felder der Versicherungen und des Detailhandels am häufigsten den Berufstitel »Data Scientist« für ihre Inserate verwenden (je 25 %).

zeigt sich, dass Organisationen dieser Felder durch sprachliche Festlegungen stärker auf ein inländisches Arbeitskräftepotenzial fokussieren.

Insgesamt präsentiert sich das untersuchte Arbeitsmarktsegment demnach zwar als feldübergreifend verteilt, weist allerdings eine erhebliche Konzentration auf Grossunternehmen in ausgewählten ökonomischen Feldern auf, die stark internationalisiert und an wenigen urbanen Standorten wie Zürich, Basel oder in der Genferseeregion lokalisiert sind. Die Differenzierung in unterschiedliche Profile erfolgt einerseits durch die jeweiligen Felder und Standorte, andererseits durch die Kombination multipler epistemischer Praktiken, disziplinärer Wissensbestände, methodischer Vorgehensweisen sowie technologischer Objekte. Wenn die Anzahl publizierter Stellenanzeigen für Data Scientists und andere datenbezogene Rollenbezeichnungen einen Gradmesser der Datafizierung darstellt, weisen die Analysen in diesem Teilkapitel auf signifikante Differenzen zwischen Feldern, Organisationen, geographischen und sprachlichen Räumen hin.

## 5.3 Repräsentationen von »Data Science« in den englischsprachigen Stellenanzeigen

Nachdem ich im vorangehenden Abschnitt die Entwicklung und Struktur der Datenwissenschaften im schweizerischen Arbeitsmarkt untersucht habe, gehe ich im Folgenden auf die Ergebnisse des Topic-Modeling-Verfahrens ein. Zunächst charakterisiere ich die Topics der englischsprachigen Stellenanzeigen allgemein und verdeutliche deren Breite, die sich trotz der Dominanz der ökonomischen Felder erkennen lässt (Kap. 5.3.1). Anschliessend gehe ich spezifisch auf drei Methoden-Topics ein, denen feldübergreifend eine zentrale Bedeutung in der Herausbildung der Datenwissenschaften zugesprochen wird (Kap. 5.3.2).

### 5.3.1 Die Vielfalt der Topics

Tabelle 5 gibt einen Überblick über die thematischen Makrostrukturen, die die englischsprachigen Stellenanzeigen prägen. Darin sind die zwanzig Topics des berechneten Modells mit den jeweils zwanzig wahrscheinlichsten Begriffen aufgeführt.<sup>15</sup> Während die Reihenfolge der Topics zufällig ist, sind die Begriffe der Topics nach ihrer Wahrscheinlichkeit geordnet. Zur Benennung der Topics wurden sowohl die Begriffswahrscheinlichkeiten als auch Dokumente, die durch die jeweiligen Topics gut beschrieben werden, herangezogen.

Dies wird exemplarisch anhand von Topic 6 nachvollziehbar: Die Begriffe *tobacco*, *philip* und *morris*, die hier zur inhaltlichen Charakterisierung des Topics verwendet werden, sind nicht unter den zwanzig häufigsten Begriffen des Topics (allerdings steht das Akronym *pmi* für »Philip Morris International«). Dennoch bilden sie das Topic inhaltlich kohärent ab, was sich manuell überprüfen lässt: Topic 6 beschreibt Stelleninserate, die zu rund 88 % von Philip Morris stammen. Während zunächst vor allem

<sup>15</sup> Aufgrund der statistischen Funktionsweise von LDA weisen nicht alle Topics dieselbe semantische Kohärenz auf und sind wegen der asymmetrischen Schätzung der Parameter verschieden gross (Kap. 4.2.5).

nicht-tabakbezogene Begriffe wahrscheinlich sind (wie *product, intern, work, develop, team, compani, futur, present, market, world*), finden sich des Weiteren auch *tobacco, cigarette, smoke-fre* oder *smoke* unter den wichtigsten Begriffen. Die inverse Verteilung der Auftretenswahrscheinlichkeiten deutet an, dass Philip Morris tabakbezogenes, d. h. feldspezifisches Vokabular zugunsten der organisationalen Selbstbeschreibung in seinen Stelleninseraten in den Hintergrund rückt.

Insgesamt fördert das Modell vier Gruppen von Topics zutage: Eine erste Gruppe von Topics umfasst Begriffe, die Datenwissenschaften primär durch Methoden bzw. epistemische Praktiken repräsentieren. Diese finden transversal Anwendung und können nicht einzelnen Feldern zugeordnet werden. Dazu zählen Project Management (T1), Internships/Science (T5), Business Reporting (T8), Data Analysis (T10), Machine Learning (T12), Digital Marketing (T14), Software Engineering (T16) sowie Business Analytics (T19). Die Topics weisen eine relativ ausgeglichene Verteilung auf, d. h., kein Feld hat einen Anteil über 50 %. Die niedrigsten Werte weisen die Topics Machine Learning (16 %), Business Analytics (19.3 %), Data Analysis (21.7 %) auf.<sup>16</sup> Die geringe Konzentration der drei methodischen Topics auf einzelne Felder deutet an, dass methodenbezogene Inhalte in Stellenanzeigen die am wenigsten feldspezifischen sind (vgl. unten).

Weiter existieren Konvolute epistemischer Praktiken in Verbindung mit einzelnen Feldern. Dabei kann zwischen Topics des Pharmafeldes und solchen mit Beziehungen zu anderen Feldern differenziert werden. Die Gruppe der pharmabezogenen Topics bilden Clinical Development (T3), Research/Biology (T7), Basel/Roche (T9) sowie Development/Laboratory (T18). T3 fokussiert vor allem klinische Studien und laborbezogene Forschungspraktiken, während T18 stärker Begriffe von Entwicklung und Produktion repräsentiert. T7 indiziert in erster Linie die Konvergenz von datengetriebener Forschung in Disziplinen wie Biologie oder Medizin (durch Begriffe wie *biolog, research, translat, medicin, bioinformat*) (Leonelli 2014, 2016; Trajanoski 2012). Schliesslich umfasst T9 Begriffe, die für Stelleninserate der Firma Roche (die häufigste Arbeitgeberin im Sample, vgl. Tabelle 3) typisch sind. Der Anteil des pharmazeutischen Feldes an diesen Inseraten ist mit Werten zwischen 72 % (T18) und 97 % (T9) sehr hoch. In geringerem Masse beschreiben die Topics ferner das akademische Feld sowie die Felder Gesundheit, Industrie und Lebensmittel, was auf eine gewisse Ähnlichkeit hinsichtlich epistemischer Praktiken und Praxisarrangements (wie Forschungslabatorien) hinweist.

Eine dritte Gruppe umfasst feldspezifische Praktiken wie Consulting Services (T4), Research/Science (T13), Tobacco/Philip Morris (T6), Recruitment (T17), Risk Management/Banking (T20), wobei die Namensgebung der Topics bereits das jeweilige Feld indiziert. Charakteristisch für diese Topics ist, dass epistemische Praktiken, Begriffe, Organisationsnamen und Standorte gemeinsam figurieren. Exemplarisch verdeutlicht dies T20: Für das Bankenfeld typische Begriffe (wie *risk, finance, bank, invest, asset, credit, wealth*) treten hier zusammen mit Standorten (*zürich, switzerland, suisse*) und Organisationsbezeichnungen (*credit, suisse, ubs*) auf. Die feldspezifischen Auftretenswahrscheinlichkeiten liegen bei diesen Topics zwischen 61.8 % (T4) und 96 % (T13).

<sup>16</sup> Es folgen Business Reporting (27.8 %), Software Engineering (28.9 %) sowie Digital Marketing (30.9 %).

*Tabelle 5: Topic-Modell der englischsprachigen Stellenanzeigen über 20 Topics mit den 20 häufigsten Begriffen (T = Topic) (n = Anzahl Dokumente)*

Project Management (T1)	Diversity/ Inclusion (T2)	Clinical Development (T3)	Consulting Services (T4)	Internships/Science (T5)	Tobacco/ Philip Morris (T6)
n=96	n=88	n=114	n=123	n=129	n=77
manag	work	clinic	servic	work	product
project	applic	studi	client	intern	intern
lead	inform	develop	consult	internship	work
team	googl	medic	team	project	pmi
strategi	experi	safeti	work	opportun	develop
across	skill	report	busi	start	team
global	opportun	review	manag	motiv	compani
ensur	employ	health	career	applic	futur
partner	status	includ	profession	program	present
abil	employe	regulatori	challeng	develop	market
drive	zurich	trial	develop	offer	world
strateg	nation	scientif	solut	industri	employ
plan	orient	statist	industri	letter	lead
develop	requir	plan	offer	date	continu
organ	recruit	activ	experi	month	requir
execut	protect	respons	opportun	environ	appli
leadership	develop	intern	project	person	global
implement	qualif	document	audit	master	potenti
respons	compani	scienc	technolog	lausann	switzerland
effect	locat	de	intern	interest	build

Research/ Biology (T7)	Business Reporting (T8)	Basel/Roche (T9)	Data Analysis (T10)
n=171	n=158	n=221	n=83
biolog	manag	basel	data
job	report	job	experi
research	requir	develop	analysi
develop	process	roch	tool
scientist	busi	switzerland	databas
drug	qualiti	research	process
scientif	support	function	sourc
experi	system	clinic	model
discoveri	master	digit	use
team	oper	patient	work
translat	analyst	opportun	knowledg
medicin	ensur	headquart	analys
high	skill	healthcar	sql
novarti	global	fact	visual
molecular	excel	basel-town	strong
analysi	improv	biomark	understand
bioinformat	provid	appli	collect
appli	respons	pharmaceut	analyst
includ	suppli	scientist	set
candid	inform	innov	integr

*Fortsetzung Tabelle 5*

Working Environment (T11)	Machine Learning (T12)	Research/Science (T13)	Digital Marketing (T14)	Working Conditions (T15)
n=173	n=150	n=227	n=139	n=29
team	learn	research	market	work
work	machin	applic	custom	benefit
peopl	scienc	scienc	digit	flexibl
product	statist	posit	compani	job
world	model	univers	sale	swisscom
creat	scientist	inform	product	offer
technolog	comput	institut	innov	data
develop	python	pleas	global	continu
build	experi	candid	develop	care
offic	r	comput	manag	addit
passion	algorithm	includ	base	system
grow	mathemat	appli	group	team
custom	appli	phd	job	compani
environ	program	group	industri	salari
compani	languag	contact	perform	programm
join	field	zurich	opportun	share
part	problem	swiss	strategi	famili
travel	deep	onlin	strong	product
us	engin	project	year	achiev
user	predict	eth	support	employ

Software Engineering (T16)	Recruitment (T17)	Development/ Lab (T18)	Business Analytics (T19)	Risk Management/ Banking (T20)
n=298	n=162	n=175	n=88	n=240
experi	experi	develop	analyt	manag
softwar	look	process	busi	risk
engin	skill	experi	data	team
develop	ag	product	solut	financi
technolog	pleas	technic	insight	bank
big	project	manufactur	advanc	work
solut	english	scientist	technolog	skill
design	posit	support	skill	support
platform	zürich	laboratori	intellig	invest
system	work	analyt	abil	asset
architectur	interest	knowledg	scienc	busi
technic	contact	work	tool	us
applic	send	perform	understand	within
servic	profession	respons	experi	challeng
cloud	german	lab	inform	offer
knowledg	switzerland	valid	drive	client
team	us	method	bi	financ
integr	year	activ	transform	function
secur	cv	evalu	design	corpor
comput	client	life	identifi	credit

Schliesslich umfasst eine vierte Gruppe von Topics stelleninseratebezogene Begriffe, die für die Analyse nicht relevant sind.<sup>17</sup> Dazu zählen Diversity/Inclusion (T2), Working Environment (T11) sowie Working Conditions (T15). Sie sind zu einem geringen Anteil in den meisten Dokumenten vorhanden, haben inhaltlich jedoch keine Bedeutung. Die Distribution über die einzelnen Felder ist deshalb eher breit (bei T2 und T11 liegen die Anteile jeweils unter 30%).<sup>18</sup> T2 beschreibt im Wesentlichen Diversity-Statements am Ende von Stellenanzeigen, was eine sprachlich-kulturelle Konvention im englischsprachigen Raum darstellt (Dobbin et al. 2015), die Organisationen in deutschsprachigen Inseraten kaum anwenden. Die Stelleninserate, die durch dieses Topic gut beschrieben werden, stammen denn auch überwiegend von Grosskonzernen (*Google, Zurich, eBay, Wipro, Biogen* oder *Siemens*) und internationalen Organisationen (*UNO, WHO*), die ihre Rekrutierungsaktivitäten primär auf internationale Arbeitsmärkte fokussieren.

### 5.3.2 Die feldübergreifende Zentralität der Methoden-Topics

Nach der Beschreibung der Topics gilt es die Ergebnisse zu interpretieren und zu plausibilisieren (DiMaggio et al. 2013: 583ff.). Die Entwicklung der relativen Häufigkeiten der Topics ist nicht wirklich sinnvoll interpretierbar, da sich visuell keine signifikanten Verschiebungen bestimmter Topics erkennen lassen (vgl. Abbildung 13 im Anhang). Anstelle der Gesamtheit der Topics untersuche ich deshalb vertieft die drei zentralen Methoden-Topics Data Analysis (T10), Machine Learning (T12) sowie Software Engineering (T16). Die durchschnittliche relative Häufigkeit – und damit die Bedeutung über alle Topics hinweg – steigt bei T16 sowie T12 über den Beobachtungszeitraum leicht an, während sie bei T10 konstant bleibt.<sup>19</sup>

Inhaltlich macht ein Vergleich der drei Topics-Kombinationen von Praktiken, Tools, Disziplinen und weiteren Kategorien erkennbar: T10 umfasst Praktiken wie *use, model, analys, understand, collect, integrate, analyz, integrat* oder *handl* sowie Instrumente wie *tool, databas, sql, warehous, access* oder *excel*, während explizite Nennungen bestimmter Felder oder Disziplinen fehlen. T12 hingegen sind oft Terme zugeordnet, die für unterschiedliche soziale (*scienc, scientist, industri*) bzw. disziplinäre Felder (wie *statist, comput, mathemat, engin, physic*) stehen. Enthalten sind auch Terme wie *learn, model, comput, program, predict, use* oder *develop*, die epistemische Praktiken andeuten, die als zentral für die Datenwissenschaften gesetzt werden. Bezogen auf Tools stechen *python, r* oder *c* [für C++] ins Auge. T16 schliesslich wird durch hohe Wahrscheinlichkeiten von informatik- und ingenieurwissenschaftlichen Begrifflichkeiten wie *softwar, engin, develop, technolog, solut, platform, system, architectur, technic, applic*,

<sup>17</sup> Maier et al. (2018: 108) und DiMaggio et al. (2013: 586) sprechen von »boilerplate topics« bzw. »boilerplate terms«, die keine inhaltliche Bedeutung haben, jedoch andere, inhaltlich relevante Topics schärfen.

<sup>18</sup> T15 tritt nur selten auf und umfasst Stellenanzeigen von zwei Unternehmen (Swisscom sowie Siemens).

<sup>19</sup> T10 weist (Spannweite  $R = 0.02$ ) die grösste Konstanz auf, während T12 ( $R = 0.037$ ) sowie T16 ( $R = 0.041$ ) stärkeren Schwankungen unterliegen. Die Spannweite  $R = x_{\max} - x_{\min}$  ist ein Streuungsmass, das sich aus der Differenz zwischen dem höchsten und dem niedrigsten Wert einer Verteilung ergibt (Diaz-Bone & Weischer 2015: 389).

*servic* oder *cloud* charakterisiert. Nachgelagert sind auch Praktiken wie *develop*, *design*, *integr*, *secur*, *comput*, *implement*, *test*, *deploy* und *distribut* sowie Tools wie *hadoop*, *java*, *spark*, *python* und *linux* enthalten.

Das Modell segmentiert demnach für Datenwissenschaften zentrale Praktiken, Disziplinen, Tools und weitere Begriffe in den drei Topics. Es umreisst damit drei methodische Praktiken, die feldunabhängig durch die Berufsgruppen der Data Scientists (T12), Data Analysts (T10) und Data Engineers (T16) repräsentiert werden. Die Benennung des Topics als »Machine Learning« deutet an, dass jenes Begriffskonvolut als zentral innerhalb der Datenwissenschaften identifiziert wird. Dadurch wird auch klar, weshalb der Begriff »Machine Learning« das Sample adäquater abbildet als die eigentlich angeführten Suchbegriffe »Data Science« oder »Big Data« (vgl. Kap. 5.2.1). Sowohl die geringe Konzentration auf einzelne Felder als auch die überdurchschnittlichen Auftretenswahrscheinlichkeiten der drei Topics verweisen darauf, dass die inserierenden Unternehmen und Organisationen Methoden über die Felder hinweg als relevant konstruieren.<sup>20</sup> Damit wird eine bestimmte Deutung der Datenwissenschaften anderen möglichen Alternativen vorgezogen und fixiert. Allerdings bleibt trotz der Bedeutung der Methoden-Topics ihr kumulierter Anteil im Vergleich zu den anderen Topic-Gruppen minoritär. Methoden sind für die Datenwissenschaften zwar wichtig, allerdings sind darüber hinaus epistemische Praktiken mit Bezug zu bestimmten, vor allem ökonomischen Feldern ebenso zentral.<sup>21</sup>

## 5.4 Repräsentationen der Datenwissenschaften in deutschsprachigen Stellenanzeigen

Nach den englischsprachigen untersuche ich nun die deutschsprachigen Stellenanzeigen. Zunächst beschreibe ich das identifizierte Topic-Modell und die Ausprägungen der einzelnen Topics, wobei englischsprachige Begriffe vor allem in den methodischen und ökonomischen Topics dominieren (Kap. 5.4.1). Danach gehe ich anhand eines Beispieldatensatzes auf die Offenheit der verwendeten Begriffe ein (Kap. 5.4.2).

### 5.4.1 Die Dominanz englischsprachiger Begriffe in deutschsprachigen Stellenanzeigen

Tabelle 6 gibt einen Gesamtüberblick über die prägenden thematischen Makrostrukturen. Darin enthalten sind die elf Topics des berechneten Modells der deutschsprachigen Stellenanzeigen mit den jeweils zwanzig wahrscheinlichsten Begriffen. Das

<sup>20</sup> Das Ergebnis deckt sich insofern mit der Curricula-Analyse, wonach Methoden transdisziplinär unterrichtet werden, während die inhaltliche Vertiefung (*domain knowledge*) in den einzelnen Disziplinen erfolgt (vgl. Kap. 9).

<sup>21</sup> Bezuglich berufspezifischer Anforderungen ist zu beobachten, dass kommunikative und soziale Kompetenzen sowie persönliche Eigenschaften wie Kreativität oder Neugier sowohl im englisch- als auch im deutschsprachigen Modell erscheinen. Sie lassen sich jedoch keinem Topic ausschliesslich zuordnen, sondern sind jeweils über mehrere verteilt. Obwohl dies auf ihre allgemeine Bedeutung über datenbezogene Berufsfelder hinaus verweist (Salvisberg 2010), haben Soft Skills im untersuchten Korpus somit nur eine geringe Relevanz.

Modell ist breiter verteilt, auch weil kein (ökonomisches) Feld dominiert wie im Falle des englischsprachigen Samples: Lediglich zwei Topics weisen einen feldspezifischen Anteil über 50 % auf, nämlich T8 Arbeitsumfeld/Swisscom<sup>22</sup> und T4 Datenwissenschaften/Forschung. T8, T2 Verantwortung/Entwicklung und T11 Stelleninserate zählen zu jenen Topics, die für Stelleninserate typisches Vokabular umfassen und nur wenige Dokumente umfassend beschreiben. T5 Schweiz – International enthält vor allem Begriffe, die das Verhältnis der Schweiz zur Welt charakterisieren.<sup>23</sup> Hier liegt die Annahme nahe, dass es sich um Inserate von Organisationen handelt, die Arbeitskräfte für Praktiken des Austauschs mit internationalen Handelspartner\*innen und anderen Akteur\*innen suchen – was bei englischsprachigen Anzeigen möglicherweise zwar enthalten ist, aber nicht explizit angegeben wird. Demgegenüber beschreibt T4 Inhalte, die für die Datenwissenschaften im akademischen Feld stehen, ohne allerdings Methoden oder epistemische Praktiken zu nennen. Neben Wissenschaft und Bildung (51 % der Dokumente in T4) sind darin ferner der öffentliche Dienst (18 %) sowie das Gesundheitswesen (14 %) vertreten.

Neben T4 segmentiert das deutschsprachige Modell jene Topics sehr ähnlich wie im Falle des englischsprachigen Samples, die epistemische Praktiken in den Stelleninseraten beschreiben: Dazu gehören T1 Big Data/Software Engineering, T3 Data Science/Machine Learning, T6 Business Analytics, T7 Datenanalyse/Reporting sowie T9 Digital Marketing. Der Umstand, dass die Topics sprachübergreifend ähnlich differenziert werden, deutet darauf hin, dass sich datenbezogene Praktiken in verschiedenen Sprachräumen etabliert haben und Anwendung finden. T1 und T3 zeichnen sich vor allem durch viele englischsprachige Terme, die Nennung spezifischer Disziplinen und Tools unter den wahrscheinlichen Begriffen aus. Analog zum englischsprachigen Modell repräsentieren die zwei Topics epistemische Praktiken, die feldübergreifend Anwendung finden (die Anteile liegen jeweils unter 30 %) und nach T7 die höchste Auftretenswahrscheinlichkeit haben. Die zwei Methoden-Topics sind demnach etwas öfter als im statistischen Durchschnitt zu erwarten, erklären aber auch im deutschsprachigen Modell lediglich 20 % der Inhalte.

T7 und T9 beinhalten – anders als beim englischsprachigen Modell – nur wenig methodisch-technisches Vokabular oder bestimmte Tools, dafür hohe Anteile an Begriffen, die in ökonomischen Feldern verwendet werden. T9 umfasst außerdem Terme, die mit der digitalen Transformation assoziiert werden, während T10 Innovation/Produktentwicklung vor allem die Entwicklung neuer Produkte und Innovation beschreibt. T7, T9 und T10 weisen alle niedrige feldspezifische Anteile auf ( $\leq 30\%$ ), während T6 vergleichsweise stark im Consulting-Feld (Muzio et al. 2011; Schmidt-Wellenburg 2013) verankert ist. T6, T7, T9 und T10 bilden zusammen die businessbezogenen Topics im Modell der deutschsprachigen Inserate. Insgesamt übertrifft die kumulierte durchschnittliche Auftretenswahrscheinlichkeit dieser für die ökonomischen Felder relevanten Topics – analog zum englischsprachigen Modell – diejenige der methodenbezogenen Topics deutlich.

<sup>22</sup> T8 wird als einziges Topic im deutschsprachigen Modell massgeblich von einem Unternehmen geprägt, da 97 % der Dokumente von der Swisscom stammen.

<sup>23</sup> Ein ähnliches Themenfeld fördert auch die Analyse der politischen Dokumente zutage (vgl. Kap. 6.3.4).

*Tabelle 6: Topic-Modell der deutschsprachigen Stellenanzeigen über 11 Topics mit den 20 häufigsten Begriffen (T = Topic) (n = Anzahl Dokumente)*

Big Data/Software Engineering (T1)	Verantwortung/ Entwicklung (T2)	Data Science/ Machine Learning (T3)	Data Science/ Forschung (T4)	Schweiz – International (T5)	Business Analytics (T6)
n=184	n=55	n=167	n=125	n=125	n=65
data	uns	data	bereich	gut	uns
big	schweiz	statist	data	kenntnis	kund
erfahr	mitarbeit	scienc	verfug	international	business
softwar	arbeit	scientist	scienc	ag	analytics
losung	verschied	analytics	informat	deutsch	consultant
engine	bereich	dat	weit	such	intelligenc
kenntnis	verantwort	r	information	bereich	sap
entwickl	sbb	neu	per	prozess	bereich
sql	vielfalt	python	gern	data	data
technologi	sich	analys	forschung	schweiz	person
kund	entwickl	mathemat	wissenschaft	aufgab	arbeit
technisch	stell	erfahr	vereinbar	fuhrrend	projekt
design	moglich	learning	team	resourc	mitarbeit
informat	wichtig	machin	erfahr	m	hast
bern	gut	method	hochschul	durchfuhr	mehr
microsoft	beweg	kenntnis	mitarbeit	erfahr	bi
datenbank	bietet	z.b	entwickl	global	technisch
team	dat	analyt	auskunft	team	senior
zurich	unternehm	modell	wirtschaft	human	losung
w	intern	w	engineering	manag	mocht

Fortsetzung Tabelle 6

Datenanalyse/ Reporting (T7)	Arbeitsumfeld/ Swisscom (T8)	Digital Marketing (T9)	Innovation/ Produkte (T10)	Stelleninserate (T11)
n=203	n=32	n=112	n=102	n=112
gut	arbeit	digital	uns	zurich
erstell	swisscom	kund	neu	ag
analys	gut	business	team	such
analyst	woch	erfahr	arbeit	aufgab
hoh	kann	data	produkt	m
analyt	beruf	analytics	bring	partn
kenntnis	uns	schweiz	zukunft	per
weiterentwickl	grosszug	management	gross	spannend
vorteil	unterstutz	m	innovativ	tätig
fahig	zusatz	neu	biet	profil
intern	famili	zurich	entwickeln	unterstutz
dat	fair	w	zusamm	bereich
business	weit	marketing	unternehm	uns
ausgepragt	moglich	digitalisier	entwickl	deutsch
management	vorteil	strateg	hast	w
reporting	wichtig	projekt	ag	freu
erfahr	profitierst	manag	herausforder	interess
person	leistung	berat	ide	unternehm
arbeitsweis	arbeitsbeding	umsetz	gemeinsam	send
auswert	gewahr	innovation	erst	bewerbungsun- terlag

#### 5.4.2 Die Offenheit der verwendeten Begriffe

Im Folgenden untersuche ich am Beispiel des bereits eingeführten Beispieldatensatzes (vgl. Anhang), wie durch offene, vieldeutige Begriffe Relationen zwischen verschiedenen Profilen und Tätigkeitsfeldern hergestellt werden, die im Sinne ambiguer Kategorien fundierend für die Datenwissenschaften sind.

Das Stelleninserat mit dem Titel »Data Scientist/Analyst (w/m) BMW Financial Services« stammt von der BMW Group Switzerland. Im Rahmen der Datenerhebung habe ich es am 1. September 2017 auf der Online-Stellenplattform jobs.ch erfasst. Bereits der Stellentitel markiert eine gewisse Unentschiedenheit, ob ein/e Data Scientist, ein/e Data Analyst oder beides gesucht wird. Dies kann eine Strategie der inserierenden Organisation sein, erst aufgrund der eingehenden Bewerbungen für eine bestimmte Berufsbezeichnung zu optieren. Da der Inseratetext weder eine Präferenz noch eine Differenz (im Sinne einer oder-Formulierung) ausdrückt, sondern an mehreren Stellen von einer »Position« bzw. einer »Persönlichkeit« spricht, artikuliert sich darin eine Offenheit über die Abgrenzung der zwei Profile bzw. Berufsgruppen, die sowohl mit unterschiedlichen epistemischen Praktiken, Objekten sowie ökonomischen Feldern verbunden werden.

Aufgrund der inserierenden Organisation habe ich das Inserat dem Transport- und Logistikfeld zugeordnet. Die Angabe »Financial Services« im Titel indiziert allerdings ein inhaltliches Praxisfeld, das eher mit Finanzdienstleistungen und damit dem Banken-/FinTech-Feld assoziiert ist. Hier bietet sich ein Rückgriff auf die identifizierten Topics an: Das Topic-Modell der deutschsprachigen Inserate beschreibt das Transport- und Logistikfeld vor allem durch die Topics 2 und 5, während jene des Banken- und Finanzdienstleistungsfeldes hohe Anteile der Topics 7 und 9 aufweisen (vgl. Tabelle 6). Das gewählte Verfahren beschreibt aber nicht nur auf der aggregierten Feldebene, sondern auch einzelne Inserate als Zusammenstellung unterschiedlicher Themen.

So setzt sich das Beispielinserat gemäss dem berechneten Topic-Modell als Kombination aus 23 % T9 Digital Marketing, 16 % T7 Datenanalyse/Reporting, 15 % T3 Data Science/Machine Learning, 12 % T11 Stelleninserate zusammen. Die restlichen sieben Topics ( $\leq 8\%$ ) machen 34 % der Inhalte aus. Der bedeutende Anteil von T9 wird insbesondere in den Abschnitten [3] und [4] deutlich: Die Position ist im »Vertriebs- und Marketingteam« angesiedelt und zählt unter anderem die Analyse und Bewertung der »Vertriebs- und Marketingperformance«, die Erstellung »komplexe[r] Vertriebs-, Markt- und Wettbewerbsanalysen« oder die Verantwortung der »Pricingstrategie« als »Aufgabengebiet« auf. T7 und T3 finden sich primär in den Abschnitten [5] und [6]: Disziplinäre und methodische Terme (»systematische Muster«, »cluster«, »Information Retrieval«, »Machine Learning«, »Naturwissenschaft«, »Statistik«, »statistische[r] Software, R, Python, SAS oder SPSS«) sind T3 zuzurechnen, während T7 eher das betriebswirtschaftliche Vokabular (»Kundensegmente [...], Trends, Kundenwerte und Handlungsbedarf (next best action)« sowie Qualifikationsanforderungen (»betriebswirtschaftliche Ausbildung (HF, FH, Uni) mit vertieften Kenntnissen in [...] Analysen und Controlling«) erfasst. T11 verweist schliesslich auf einen vergleichsweise hohen Anteil stelleninseratetypischen Vokabulars (Abschnitte [3], [6] und [7]). Über den gesamten Inseratetext dominieren also betriebswirtschaftliche, marketingbezogene Inhalte, während das explizit methodenbezogene Vokabular dieses komplementiert.

Das Topic-Modell gibt keine Antwort darauf, ob in diesem Inserat eher ein/e Data Scientist oder ein/e Data Analyst gesucht wird, sondern identifiziert die Inhalte, die durch diese Topics beschrieben werden, als quasi gleichwertig. Damit löst die Analyse des Inseratetextes auch die Unklarheit bezüglich der Kategorisierung der Stellenbezeichnung nicht auf. Obwohl die Differenz *Data Scientist* oder *Data Analyst* offengelassen wird, weisen Stellentitel und Text dennoch eine inhaltliche Kohärenz auf, die durch das deutschsprachige Topic-Modell und die Verteilung der Inhalte nachvoll-

ziehbar wird. Die Offenheit bezüglich der in diesem Inserat verwendeten Kategorien ist jedoch konstitutiv, da sie multiple Handlungs-, d. h. in diesem Fall Rekrutierungs- optionen bereithält. Ambiguität (Eisenberg 1984; Leitch & Davenport 2007) in Stellen- anzeigen als organisationalen Praktiken wird so zu einem zentralen Merkmal in der Herausbildung der Datenwissenschaften.

## 5.5 Diskussion

Das Kapitel hat untersucht, wie Akteur\*innen in unterschiedlichen Feldern durch die Publikation von Stellenanzeigen multiple Bedeutungen der Datenwissenschaften konstruieren. Ich habe zunächst gefragt, wie das Segment der Datenwissenschaften im Arbeitsmarkt strukturiert ist und welche Organisationen in welchen Feldern Stelleninserate für Data Scientists publizieren. Danach erschloss die Analyse die unterschiedlichen Profile von Data Scientists, die in den Inseraten konstruiert werden. Schliesslich interessierte sich das Kapitel dafür, welche übergeordneten Deutungen der Datenwissenschaften in den Stelleninseraten entworfen und fixiert werden.

Die Ergebnisse der deskriptiven Analyse verweisen einerseits auf die zunehmende Bedeutung der Datenwissenschaften während des Untersuchungszeitraums. Andererseits indizieren sie eine ausgeprägte Konzentration der Stelleninserate auf bestimmte Grossunternehmen in ökonomischen Feldern, die in urbanen Standorten wie Zürich und Basel lokalisiert sind. Der hohe Anteil englischsprachiger Inserate im Sample kann als ein Indikator für die starke Internationalisierung des Arbeitsmarkt- segments interpretiert werden. Insbesondere Organisationen in den ökonomischen Feldern Arzneimittelherstellung, Banken, IT/Software und Consulting sowie im akademischen Feld publizieren besonders viele Stelleninserate für Data Scientists. Dies kann als Indiz dafür gelesen werden, dass sie stärker als andere datafiziert sind. Es zeigen sich zudem signifikante Unterschiede zu Organisationen in anderen Feldern, Landesteilen und Sprachregionen, die noch weniger durch Datenpraktiken durchdrungen sind. Insofern manifestieren sich erhebliche Differenzen in der Diffusion und Adoption datenwissenschaftlicher Expertise und Praktiken in den einzelnen Feldern.

Über die Titel der Stelleninserate können verschiedene Gruppen von datenbezogenen Berufsbezeichnungen differenziert werden, die im Wesentlichen jenen in der Literatur geforderten und identifizierten Profilen von Data Scientists entsprechen. Über die Rollenbezeichnungen hinaus lassen sich diese einerseits den genannten (ökonomischen) Feldern und Standorten zuweisen. Andererseits sind die Profile durch distinkte Begriffskonvolute charakterisiert, die epistemische Praktiken, disziplinäre Wissensbestände, methodische Vorgehensweisen und Objekte miteinander verknüpfen. »Data Science« und »Big Data« als der Analyse zugrundeliegende Suchterme erweisen sich damit als inklusive Begriffe, durch die multiple Alternativen formuliert werden. Solche Praktiken von Begriffsarbeit eröffnen den inserierenden Organisationen neue Möglichkeitsräume und erweitern somit deren Handlungs- bzw. Rekrutierungsoptionen.

Gegenüber der strukturellen Konzentration der Stelleninserate fördert das Topic- Modeling-Verfahren auf der inhaltlichen Ebene sowohl für das englisch- als auch das deutschsprachige Sample die Multidimensionalität der Datenwissenschaften zutage. Die computergestützte Textanalyse der Stellenanzeigen indiziert eine grosse Vielfalt

unterschiedlicher Konnotationen, durch die die Datenwissenschaften in den Topics repräsentiert werden. Dies sind zum einen Topics, die feldübergreifend wissenschaftliche Methoden und epistemische Praktiken beschreiben. Zum anderen lassen sich Topics differenzieren, die für Wissensinhalte, Begriffe und Objekte in bestimmten Feldern, namentlich im pharmazeutischen, im Banken- sowie im akademischen Feld stehen. Die Heterogenität der identifizierten Topics bestätigt insofern die theoretische Annahme, dass sich – trotz der Dominanz der ökonomischen Felder hinsichtlich der Struktur des Arbeitsmarktsegments – die Datenwissenschaften weder auf ein soziales Feld noch auf eine Disziplin reduzieren lassen.

Insofern ist die Prävalenz pharmabezogener Topics, die im englischsprachigen Modell einen gewichtigen Anteil der Inhalte prägen, erklärbungsbedürftig: Deren Auftretenswahrscheinlichkeiten schwanken in der Untersuchungsperiode zwar relativ stark, weisen jedoch eine ähnliche Spannweite auf wie die Methoden-Topics. Die Relevanz der Topics bemisst sich in der hohen Anzahl von Stellenanzeigen des Pharmafeldes, das über einen Viertel des englischsprachigen Samples verantwortet.<sup>24</sup> Zwar werden die Naturwissenschaften bisweilen als wichtige Bezugsdisziplinen der Datenwissenschaften genannt, allerdings deutlich seltener als Statistik, Mathematik und Informatik. Gerade laborbezogene epistemische Praktiken finden in der Diskussion um die notwendigen Kompetenzen für Datenwissenschaftler\*innen kaum je Erwähnung. Möglicherweise spiegelt das Ergebnis die datengetriebene Transformation der für die Pharmaindustrie relevanten epistemischen Felder wider (Leonelli 2016; Trajanoski 2012), was sowohl hinsichtlich der ökonomischen Wertschöpfung als auch der kollektiven Selbstbeschreibung der Schweiz von elementarer Bedeutung ist.

In den zentralen methodischen Topics werden Praktiken, Tools und Disziplinen, die in der Diskussion gemeinhin mit Datenwissenschaften assoziiert werden, in den sprachgetrennt berechneten Modellen separat segmentiert. Die Kombination disziplinen- und feldspezifischer Begriffe, Methoden und Tools in diesen Topics verweist darauf, dass sich in der Diskussion bestimmte Kategorien etabliert haben, die in Stelleninseraten feldübergreifend und sprachunabhängig Verwendung finden. Sowohl die Präsenz englischsprachiger Rollenbezeichnungen als auch die Frequenzen englischsprachiger Terme in den Topics des deutschsprachigen Modells machen klar, dass es sich dabei um englischsprachige Kategorien handelt, die von ihrem Popularisierungskontext im Feld der US-amerikanischen Technologiekonzerne in diverse ökonomische und soziale Felder diffundiert sind. Die Analyse der Stelleninserate verdeutlicht auch, dass Methoden, Analysetools oder Qualifikationsanforderungen oft austauschbar sind und in nicht-hierarchisierenden Aufzählungen angeführt werden. Solche Listen rücken die unklare Verortung der Datenwissenschaften in den Vordergrund: Da für viele Organisationen nach wie vor Unsicherheit über die zentralen Begriffe besteht, signalisiert das Changieren zwischen verschiedenen Vorgehensweisen, Werkzeugen oder Bildungszertifikaten Inklosivität methodischer Expertisen sowie disziplinärer Wissensbestände. Die in den Stelleninseraten zu beobachtenden Aufzählungen äquivalenter Alternativen operieren in diesem Sinne als Praktiken der Grenzüberschreitung, die darauf abzielen, möglichst viele Wissensbereiche und Expertisen zu inkludieren.

24 Das Ergebnis könnte auch eine Folge der Suchstrategie mit den Begriffen »data« und »scientist« sein, indem der Suchalgorithmus der Jobplattform die zwei Begriffe mit Inseraten der Pharmaindustrie assoziierte und insofern zu einer Überrepräsentation dieses Feldes führte.

dieren und den Gegenstand flexibel, adaptiv und offen zu halten. Damit markiert sich eine erste Differenz zum akademischen Feld, in dem die Synchronizität divergierender Praktiken der Grenzziehung und Grenzüberschreitung fundamental in die Herausbildung der universitären Curricula eingeschrieben ist (vgl. Kap. 9).

Trotz der Bedeutung jener Topics, die Datenwissenschaften als Kombinationen methodenbezogener Begriffe und Praktiken repräsentieren, bleibt deren kumulierter Anteil in beiden Modellen in Relation zu anderen Topics minoritär. In der Verknüpfung von methodischen Praktiken und ökonomischen Wissensinhalten manifestiert sich die Priorität der oft eingeforderten Anwendungskomponente. Dies zeigt sich ferner in der Unabhängigkeit der Methoden von jenen Topics, die für das institutionelle akademische Feld stehen, ohne bestimmte Disziplinen oder Methoden zu beschreiben. Somit ergibt sich eine weitere Differenz zum akademischen Feld, wo in den Curricula die methodische Ausbildung im Kernbereich durch eine *domain knowledge* in Wahlfächern ergänzt wird. Im Korpus der Stelleninserate bilden hingegen die feldspezifischen Inhalte der *domains* den Kern, während die methodischen Kompetenzen diesen komplementieren.

Insgesamt setzen Stellenanzeigen eine Vielzahl unterschiedlicher Organisationen, Felder, Praktiken, Objekte und Begriffe miteinander in Beziehung, ohne die jeweiligen Relationen zwischen diesen zu definieren. Vielmehr sind die verwendeten Begriffe feld- und sprachübergreifend durch Anpassungsfähigkeit und Mehrdeutigkeit charakterisiert, was den Akteur\*innen affirmative Bezugnahmen erlaubt, wobei die Bedeutung der jeweiligen Begriffe von Inserat zu Inserat, von Feld zu Feld differieren kann. Die Offenheit und Ambiguität der verwendeten Begriffe ist konstitutiv für das untersuchte Segment des schweizerischen Arbeitsmarktes und fundiert insofern die Emergenz der Datenwissenschaften als ein Wissensgebiet, das sprachliche Grenzziehungen transzendieren und verschiedene etablierte Felder miteinander verknüpfen kann.