

Digitalisierung des Gesundheitssektors: Analyse der digitalen Kompetenzen von Ärzten

Marcel Herold^{1*}

¹ Technische Hochschule Wildau, Deutschland

* Korrespondenz: marcel.herold@th-wildau.de

Kurzfassung. Der Fachkräftemangel im Gesundheitssektor beschleunigt die Digitalisierung, wodurch die Anforderungen an das Personal steigen. Die zunehmenden technischen Möglichkeiten verändern die notwendigen Kompetenzen an die Ärzte in Richtung Digitalisierung. Daher sollten im Personal-Recruitment digitale Kompetenzen bereits in den Stellenausschreibungen berücksichtigt werden, um die zukünftigen Herausforderungen bewältigen zu können. Dieser Beitrag untersucht, inwieweit a) digitale Kompetenzen bereits in den am Markt befindlichen Stellenprofilen von Ärzten Eingang gefunden haben und b) ob diese durch einen datengetriebenen methodischen Ansatz sinnvoll extrahiert werden können. Dabei werden 1707 Stellenausschreibungen mit der Latent Semantic Analysis (LSA) ausgewertet. Die unterschiedlichen methodischen Ansätze innerhalb der LSA zeigen, dass kaum ein Fokus auf digitale Kompetenzen im Gesundheitssektor bei Stellenausschreibungen besteht.

1. Einleitung

Der Fachkräftemangel in Deutschland und insbesondere im Gesundheitssektor ist allgegenwärtig und wird sich durch den demografischen Wandel noch verstärken [1]. Gleichzeitig nimmt die Digitalisierung in dieser Branche zu: Eine Studie hat z.B. dargelegt, dass Technologie und Digitalisierung positive Auswirkungen auf den Krankheitsverlauf von Patienten haben [2]. Darüber hinaus kann Telemedizin auch die Behandlung von Patienten in ländlichen Gebieten sicherstellen [3]. Außerdem konnte in einigen Fällen eine höhere Therapietreue beim Verwenden von mobilen Geräten festgestellt werden [4].

Durch diese neuen technischen Möglichkeiten verändern sich die Anforderungen an das Personal [5]. Einerseits wird die Angst diskutiert, dass Arbeitsplätze verloren gehen, andererseits müssen bestehende Mitarbeiter neue Kompetenzen trainieren, um die Technologie bedienen zu können [6]. Dabei ist die Personalabteilung gefordert, die Änderungen der Anforderungen in den Stellenprofilen durch die Digitalisierung auch im Personalrecruitment, z.B. im Rahmen der Stellenausschreibungen, zu berücksichtigen [7]. So können die Herausforderungen der Digitalisierung im Gesundheitssektor noch stärker als bisher bewältigt werden.

Ziel dieses Beitrages ist daher zu untersuchen, inwieweit a) digitale Kompetenzen bereits in den am Markt befindlichen Stellenprofilen von Ärzten Eingang gefunden haben und b) ob diese durch einen datengetriebenen methodischen Ansatz sinnvoll extrahiert werden können.

2. Methode

Um diese Forschungsfragen zu beantworten, werden Stellenanzeigen im Gesundheitssektor hinblickend auf die ausgeschriebenen Kompetenzen analysiert. Die zu analysierenden Daten

wurden mit Hilfe von Webscraping von einem großen deutschen Jobportal heruntergeladen. Es wurde ein Suchstring entwickelt, um vergleichbare Stellenanzeigen von der Berufsgruppe Arzt / Ärztin zu erhalten, der verschiedene Facharztausbildungen reflektierte: „Arzt |Ärzt |Medi |Vete |Huma |Intern |Epid |Psyc |Physi“. Dabei soll mind. einer der Begriff im Jobtitel enthalten sein („|“ = Oder), unabhängig von der Platzierung am Wortanfang, -mitte oder Ende (nicht Case-Sensitiv). Somit ergab sich die Anzahl der Stellenanzeigen der Berufsgruppe „Ärzte“ mit 1.722, welche auf 1.707 reduziert werden musste, da 15 Anzeigen keine Qualifikationsanforderungen beinhalteten.

Dieser reduzierte Datensatz wurde nach Miner und Kollegen [8] vorverarbeitet. Das sog. Preprocessing beinhaltet die Schritte Großbuchstaben in Kleinbuchstaben umzuwandeln, den Text in Worte zu zerlegen (Tokenisierung), typische sprachabhängige Stoppwörter zu entfernen, kontextabhängige Tokens zu filtern, Wörter auf Ihre Grundform zurückzuführen (stemmen) und abschließend alle Begriffe nach deren Wichtigkeit im gesamten Datensatz zu gewichten. Bei dieser Analyse wurden nur die Qualifikationen der Stellenausschreibungen analysiert. Insgesamt betrug die Anzahl der Wörter 8.662.

In der Forschung zum Kompetenz-orientierten Personalmanagement gibt es verschiedene methodische Ansätze, um textbasierte Daten mit Hilfe von Text Mining und Natural Language Processing zu analysieren. In Abbildung 1 sind die unterschiedlichen Ansätze von Sidorova und Kollegen [9], Chung und Chen [10] nach Evangelopoulos [11] und Pejic-Bach und Kollegen [12] nach Jimenez-Marquez et al [13] dargestellt, welche in dieser Arbeit unter Nutzung von PYTHON mit den Paketen NLTK und SKLEARN vereinheitlichend angewendet werden. Diese werden in weiterer Folge allgemein beschrieben, bevor unter Punkt 3, Ergebnisse, ihre Anwendung auf die Forschungsfrage dargestellt wird.

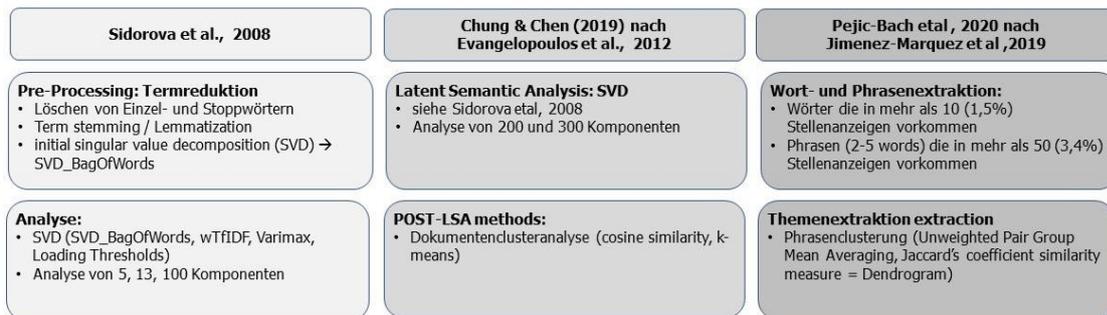


Abbildung 1. Methodisches Framework

Sidorova und Kollegen [9] analysieren 1.615 Forschungszusammenfassungen aus drei Fachzeitschriften, die zwischen 1985 und 2006 veröffentlicht wurden. Zuerst führten diese eine Term-Reduktion durch Löschung von Einzelwörtern und Stopwörtern durch. Zusätzlich wurde ein „term stemming“ angewendet, so dass noch 3.172 gestemmte Wörter zur Analyse zur Verfügung stehen. Basierend auf dieser Wortliste führten die Autoren eine „initial singular value decomposition“ (iSVD) mit 100 Komponenten aus. Darauf aufbauend wurde eine erweiterte SVD mit 5, 13 und 100 Komponenten, einer Matrix Rotation und dynamischer Komponentenladung angewendet, um die Terme zu identifizieren, welche die größte Varianz erklären.

Chung und Chen [10] haben mittels Text Mining 841 Stellenanzeigen für „HR Manager“ hinblickend auf die notwendigen Kompetenzen analysiert. Dabei haben sie auf das „Text Mining research framework“ von Evangelopoulos und Kollegen [11] zurückgegriffen. Dieses Vorgehen beinhaltet ebenfalls zuerst eine Term-Reduktion und eine iSVD mit 200 und 300 Komponenten, ergänzt das Methodenframework aber um eine Clusteranalyse auf Basis der Cosi-

nus-Distanzen der Dokumente. Sie konnten insgesamt 5 Cluster in den Stellenanzeigen extrahieren und zeigten zudem eine deskriptive Datenanalyse (Erfahrungen, Ausbildung und Sprache).

Pejic-Bach und Kollegen [12] haben genauso wie Chung und Chen [10] Stellenanzeigen mit dem Fokus auf „Industry 4.0“ basierend auf Text Mining analysiert. Zur Analyse der Stellenanzeigen haben sie den Ansatz von Jimenez-Marquez und Kollegen [3] verwendet. Im ersten Schritt haben sie statt des Stemming die Lemmatisierung und gleichlautend die Term-Reduktion durchgeführt. Im nächsten Schritt haben sie die häufigsten Wörter, die in mehr als zehn Stellenanzeigen vorkommen, extrahiert. Danach wurden die häufigsten Phrasen mit maximal 5 Wörtern, die in mehr als 50 Stellenanzeigen vorhanden waren, extrahiert. Im letzten Schritt wurde eine Clusteranalyse für die extrahierten Phrasen durchgeführt. Sie konnten acht Cluster identifizieren.

3. Ergebnisse

Gleichlautend für alle Analysen wurden, wie oben erwähnt, die 8.662 Worte der Qualifikationsanforderungen von 1.707 Stellenanzeigen verwendet. Vereinheitlichend wurde die Lemmatisierung der Worte genutzt, um diese auf Ihre Wortstämme zurückzuführen, da diese dem Stemming mit einem festen Algorithmus überlegen ist [14].

Bei der Vorgehensweise von Sidorova und Kollegen [9] wurden 3.367 lemmatisierte Wörter analysiert, da 4709 Einzel- und Stoppwörter gelöscht wurden. Die iSVD mit 100 Komponenten klärt mit 1020 Wörtern 56,9 Prozent der Dokumentenvarianz auf. Bei der erweiterten SVD wurde wie vorgeschlagen 5, 13 und 100 Komponenten für die gewichtete „Document term matrix“ verwendet. Dabei können 5 Komponenten 8,9 Prozent der Varianz mit 560 Einzelwörtern erklären. 13 Komponenten können 16,5 Prozent der Varianz mit 416 Einzelwörtern erklären und 100 Komponenten können 49,8 Prozent der Varianz mit 149 Wörtern erklären. Keine der getesteten Komponenten-Anzahl zeigte eine oder mehrere Komponenten mit eindeutigem technischem Bezug, dagegen wurden Wörter, wie „EDV“ oder „MS Office“ als generische oder „Radiologie“ und „Neurologie“ als spezifische Begriffe für digitale Technologien verteilt in verschiedenen Komponenten identifiziert (siehe Tabelle 1), „Telemedizin“ im konkreten tauchte als Begriff dagegen nicht auf.

Tabelle 1. Übersicht Komponenten mit generischen und spezifischen technischen Begriffen

Autor	Komp .	Komp. mit technischen Begriffen	Beispiele generische technische Begriffe	Beispiele spezifische technische Begriffe
Sidorova et al [9]	5	5	it, microsoft, sap, edv, online, office, digitale, technische	neurologie, endoskopie, arzt-informationssystem, kardiologie
Sidorova et al [9]	13	12	sap, powerpoint, office, outlook, software, google	endoskopie, kardiologie, radiologie
Sidorova et al [9]	100	51	digital, edv, ms, teams, it	radiologie, neurologie
Evangelopoulos et al [11]	200	72	ms teams, sap, edv, office, pc, word, technische, it	neurologie, radiologie
Evangelopoulos et al [11]	300	95	edv, it, technisch, ms tams, office, word, excel, pc	ekg, neurologisch, radiologie, kardiologie

1.636 Begriffen und das Kleinste mit 51 (non-unique) Begriffen. Auch hier wurden bei vier von den sieben Clustern generische Begriffe der Digitalisierung wie „EDV“, „Outlook“ oder „PC“ und als spezifische Begriffe „ultraschall“ oder „ekg“ identifiziert. Somit erzeugte auch die Vergrößerung der initialen Komponenten-Anzahl am Ende keine eindeutige technische Komponente.

Nach Pejic-Bach und Kollegen [12] wurde im ersten Schritt alle Wörter identifiziert, die in mind. 10 Stellenanzeigen vorhanden sind. Dies entspricht eine Quote von ca. 1,5 Prozent. So konnten insgesamt 379 Worte identifiziert werden. Es wurden beispielsweise Begriffe wie „MS“, „Office“, „EDV“, „it“, „media“, „excel“, „sap“, „pc“ oder „outlook“ herausgefiltert, welche im Zusammenhang mit den digitalen Kompetenzen stehen. Im zweiten Schritt wurden Wortphrasen mit 2-5 Wörtern, welche in mind. 50 Stellenanzeigen auftauchen, gesucht (dies entspricht etwa 3.4%). Insgesamt erfüllten 51 Phrasen diese Bedingungen. Im letzten Schritt wurde basierend auf den Phrasen eine Clusteranalyse durchgeführt. Dabei wurden erneut 7 Cluster entdeckt, wobei diesmal nur ein Cluster mit den 3 Phrasen a) „umgang ms office“, b) „ms office“ und c) „umgang ms“ den digitalen Kompetenzen hinzugerechnet werden konnte.

4. Diskussion & Ausblick

Durch die zunehmende Digitalisierung im Gesundheitssektor [2, 3, 4,] und die steigenden technischen Anforderungen an die Ärzte können die notwendigen Kompetenzen nicht eindeutig aus den Stellenanzeigen herausgefiltert werden. Es sind generisch technische Begriffe in den Stellenausschreibungen enthalten, aber kaum spezifische. Die verwendeten Analysen bekräftigen, dass die Extraktion von technologisch geprägten Komponenten auf Basis der analysierten Stellenprofile schwierig ist. Bei dem Ansatz von Sidorova und Kollegen [9] konnte bei 5, 13 und 100 Komponenten Begriffe (Tabelle 1) gefunden werden, die in Verbindung mit den digitalen Kompetenzen stehen. Bei dem Ansatz von Evangelopoulos und Kollegen [11] waren Begriffe mit Bezug zu digitalen Kompetenzen über vier Cluster verteilt. Bei dem Vorgehen von Pejic-Bach und Kollegen [12] konnten eine geringe Anzahl an Einzelworten und Wortphrasen herausgefiltert werden, welche in Bezug zu den digitalen Kompetenzen stehen. Durch diese Ergebnisse kann die Forschungsfrage nicht eindeutig beantwortet werden.

Ein weiterführender Ansatz zur Identifizierung von digitalen Kompetenzen im Gesundheitssektor ist die Optimierung des datengetriebenen Ansatzes. Dazu soll als erstes die Log Entropie statt die TF-IDF verwendet werden. Als zweites soll die LSA durch die Latent Dirichlet Allokation ersetzt werden. Als Drittes soll auf Basis der Taxonomie für Kompetenzen in diesem Sektor [15] eine eigene „Bag of Word“-Liste mit den digitalen Kompetenzen erstellt werden, um die Stellenausschreibungen zu analysiert. Bei dieser Analyse könnten dann auch die Pflgestellten und die Arztstellen miteinander verglichen werden.

Datenverfügbarkeit

Die Daten können auf Nachfrage zur Verfügung gestellt werden.

Interessenskonflikte

Der Autor erklärt, dass er keinen Interessenkonflikt hat.

Literaturverzeichnis

1. Reiff, E., Gade, C., & Böhlich, S. (2020): Handling the shortage of nurses in Germany: Opportunities and challenges of recruiting nursing staff from abroad, IUBH Discussion

2. Hansen, A. et al. (2019): Perception of the Progressing Digitization and Transformation of the German Health Care System Among Experts and the Public: Mixed Methods Study. *JMIR Public Health and Surveillance*, 5/4: e14689.
3. Davis, A. M., et al. (2016): Treating rural paediatric obesity through telemedicine vs. telephone: Outcomes from a cluster randomized controlled trial. *Journal of Telemedicine and Telecare*, 22/2: 86–95.
4. Frias, J. et al (2017): Effectiveness of Digital Medicines to Improve Clinical Outcomes in Patients with Uncontrolled Hypertension and Type 2 Diabetes: Prospective, Open-Label, Cluster-Randomized Pilot Clinical Trial. *Journal of Medical Internet Research*, 19/7: e246.
5. Veit, K., Wessels, M., & Deiters, W. (2019): Gesundheitsdaten und Digitalisierung – Neue Anforderungen an den Umgang mit Daten im Gesundheitswesen. In M. A. Pfannstiel, P. Da-Cruz, & H. Mehlich (Hrsg.), *Digitale Transformation von Dienstleistungen im Gesundheitswesen VI* (S. 19–33). Springer Fachmedien Wiesbaden.
6. Spencer, D. A. (2018): Fear and hope in an age of mass automation: Debating the future of work. *New Technology, Work and Employment*, 33/1: 1–12.
7. Karwehl, L. J., & Kauffeld, S. (2021): Traditional and new ways in competence management: Application of HR analytics in competence management. Gruppe. Interaktion. Organisation. Zeitschrift Für Angewandte Organisationspsychologie (GIO), 52/1: 7–24.
8. Miner, G. D., Delen, D., Elder, J., Fast, A., Hill, T., & Nisbet, R. A. (2012). *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. Elsevier.
9. Sidorova, et al. (2008): Uncovering the Intellectual Core of the Information Systems Discipline. *MIS Quarterly*, 32/3: 467.
10. Chung, C.-H., & Chen, L.-J. (2019): Text mining for human resources competencies: Taiwan example. *European Journal of Training and Development*, 45(6/7): 588–602.
11. Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012): Latent Semantic Analysis: Five methodological recommendations. *European Journal of Information Systems*, 21/1, 70–86.
12. Pejic-Bach, M., et al. (2020): Text mining of industry 4.0 job advertisements. *International Journal of Information Management*, 50: 416–431.
13. Jimenez-Marquez, J. L., et al. (2019): Towards a big data framework for analyzing social media content. *International Journal of Information Management*, 44: 1–12.
14. Balakrishnan, V., & Ethel, L.-Y. (2014): Stemming and Lemmatization: A Comparison of Retrieval Performances. *Lecture Notes on Software Engineering*, 2(3): 262–267.
15. Englander, R., et al. (2013): Toward a Common Taxonomy of Competency Domains for the Health Professions and Competencies for Physicians: *Academic Medicine*, 88/8: 1088–1094.